Lecture II
Introduction to complex networks

Santo Fortunato

INSTITUTE
FOR SCIENTIFIC INTERCHANGE
FOUNDATION

Plan of the course

I. Networks: definitions, characteristics, basic concepts in graph theory

II. Real world networks: basic properties. Models I

III. Models II

IV. Community structure I

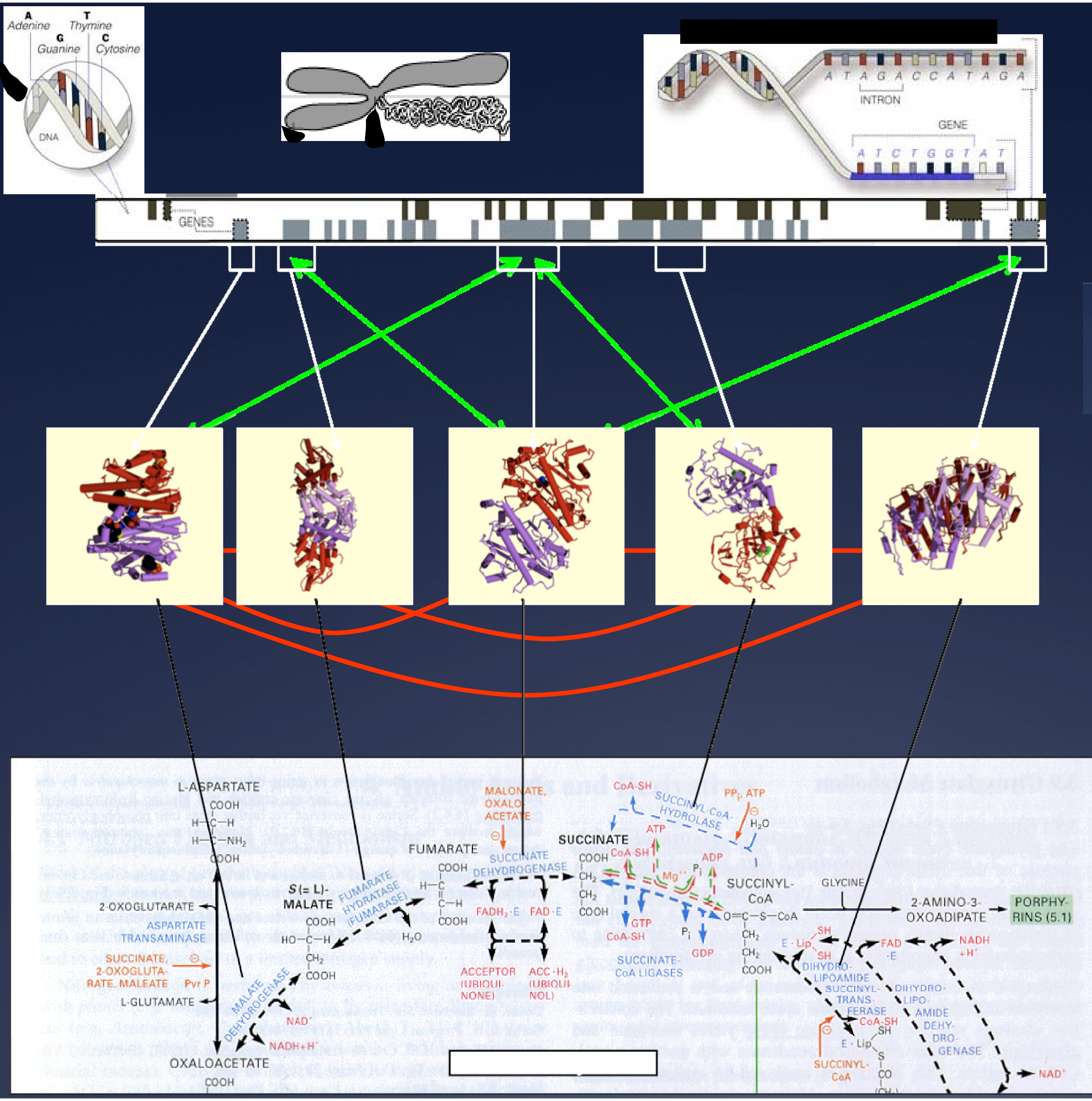
V. Community structure II

VI. Dynamic processes in networks

Two main classes

Natural systems:
Biological networks: genes, proteins…
Foodwebs
Social networks

Infrastructure networks:
Virtual: web, email, P2P
Physical: Internet, power grids, transport…



Genome

Protein-gene interactions

Proteome

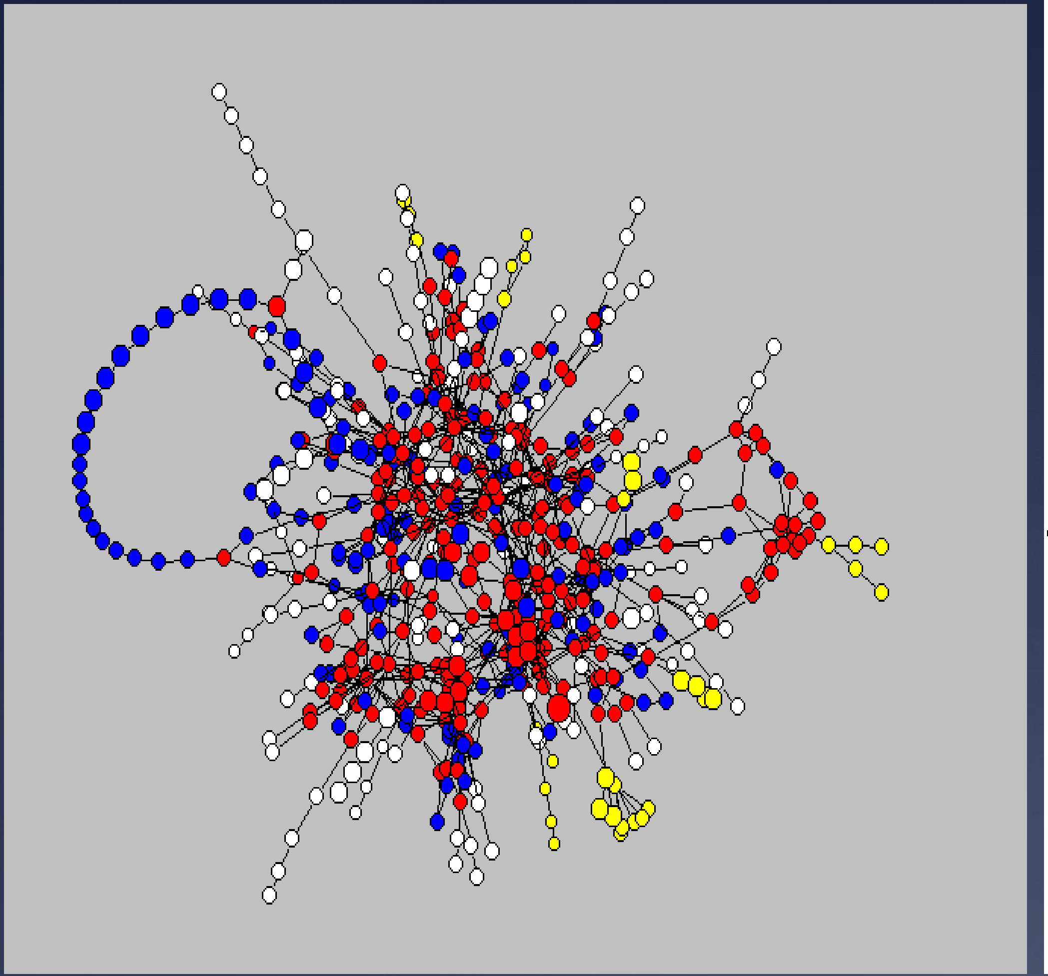
Protein-protein interactions

Metabolism

Bio-chemical reactions

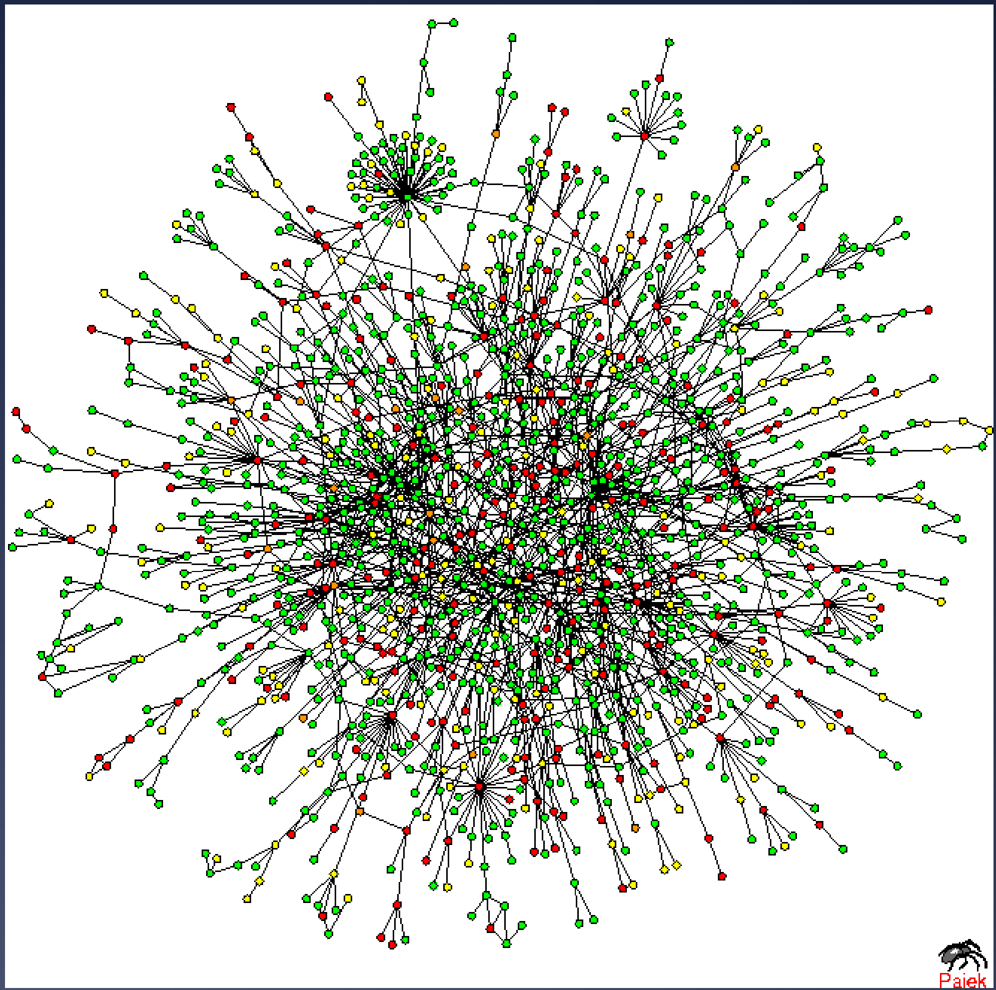
Metabolic Networks

Vertices: metabolites
Edges: chemical reactions



Protein Interactions

Vertices: proteins
Edges: interactions



Scientific collaboration networks

Vertices: scientists

Edges: co-authored papers

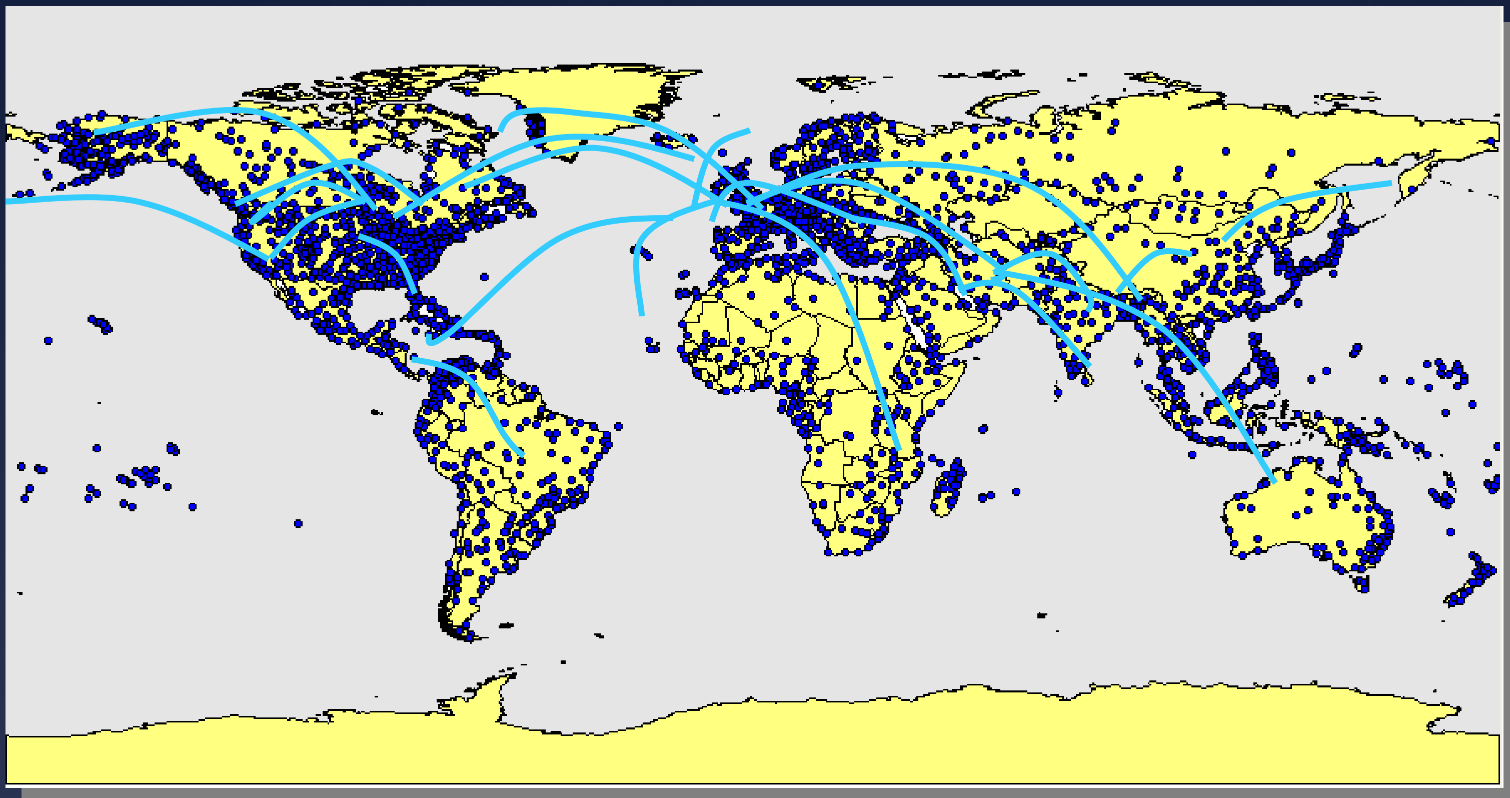
Weights: depending on
- number of co-authored papers
- number of authors of each paper
- number of citations…

Actors collaboration networks

Vertices: actors

Edges: co-starred movies

World airport network



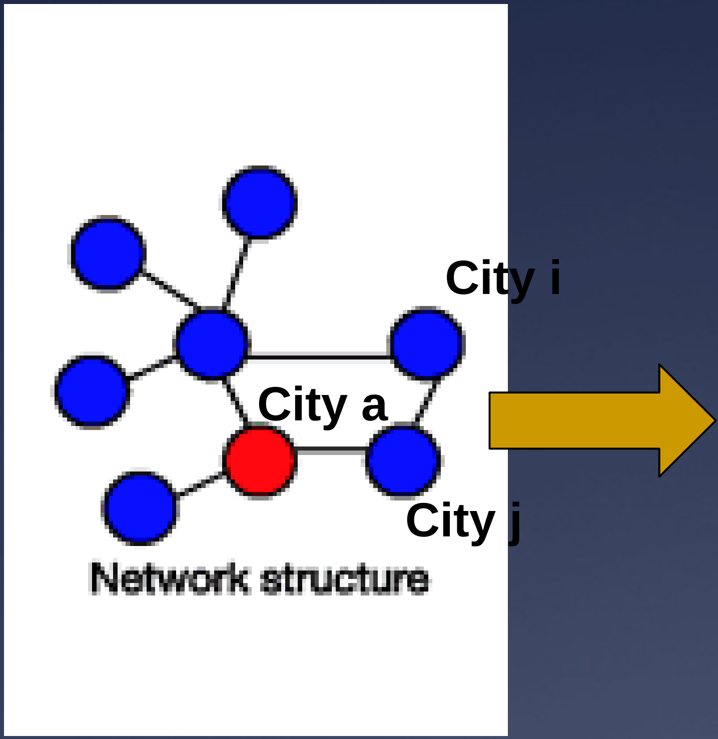
complete IATA database

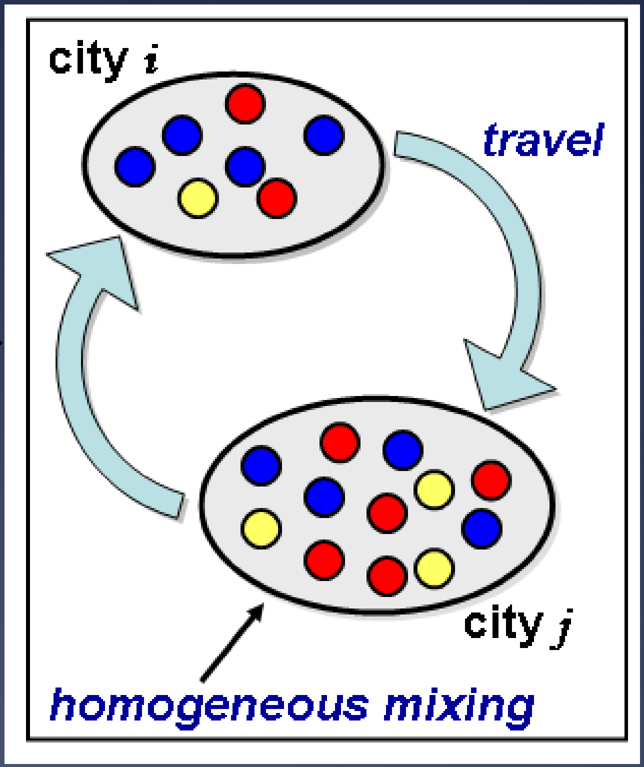
- ***V* = *3100* airports**
- ***E* = *17182* weighted edges**
- **w_{ij} #seats / (time scale)**

> 99% of total traffic

Meta-population networks

Each vertex: internal structure
Edges: transport/traffic





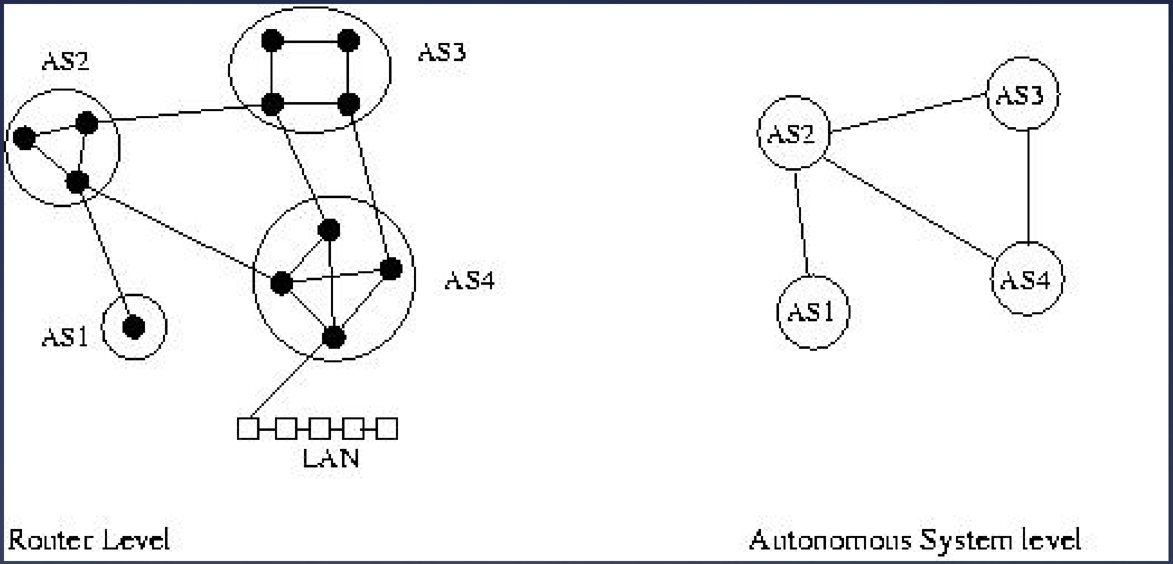
Internet

- Computers (routers)
- Satellites
- Modems
- Phone cables
- Optic fibers
- EM waves

Internet

Graph representation

different granularities



Internet mapping

- continuously evolving and growing
- intrinsic heterogeneity
- self-organizing

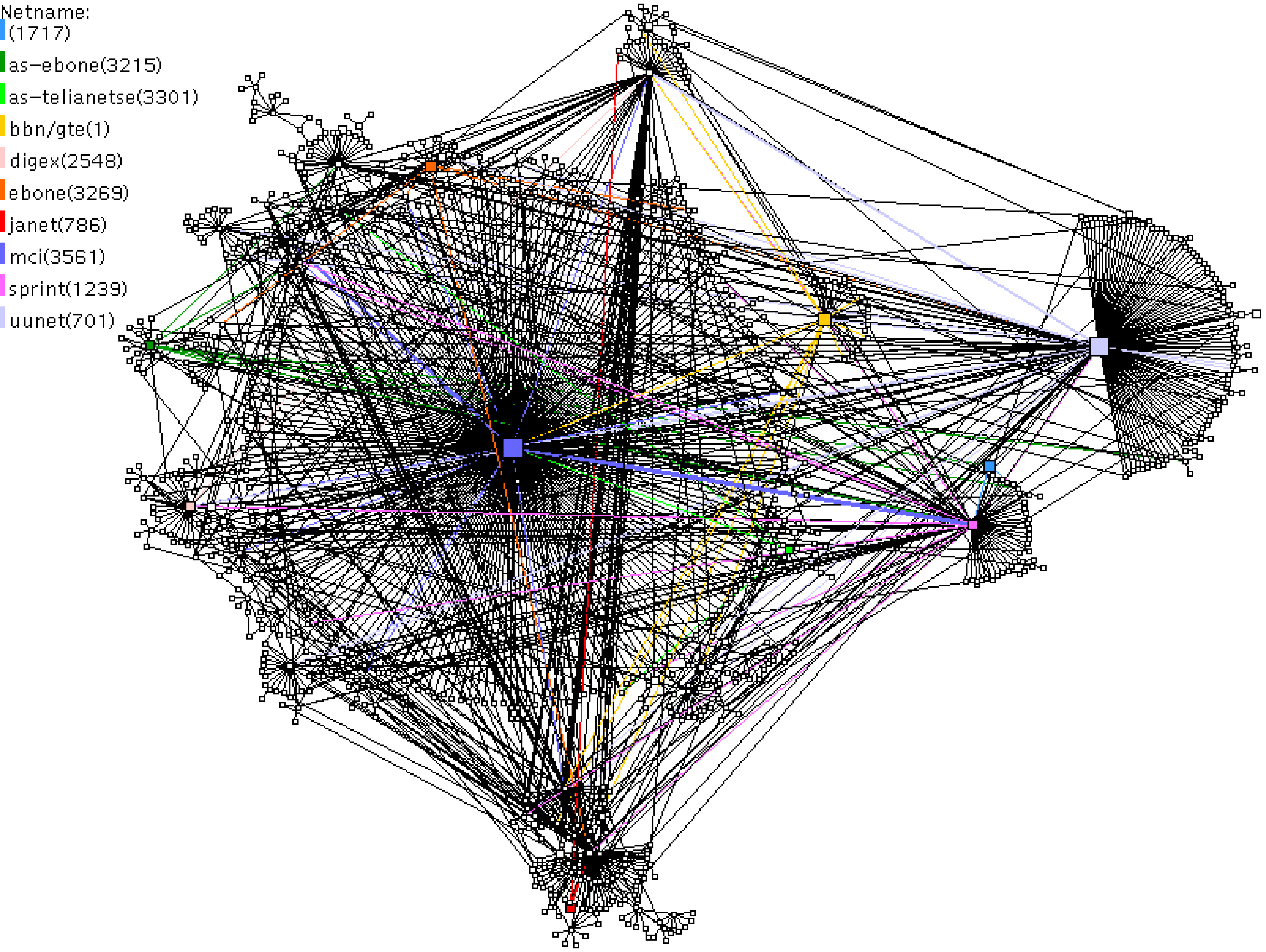
Largely unknown topology/properties

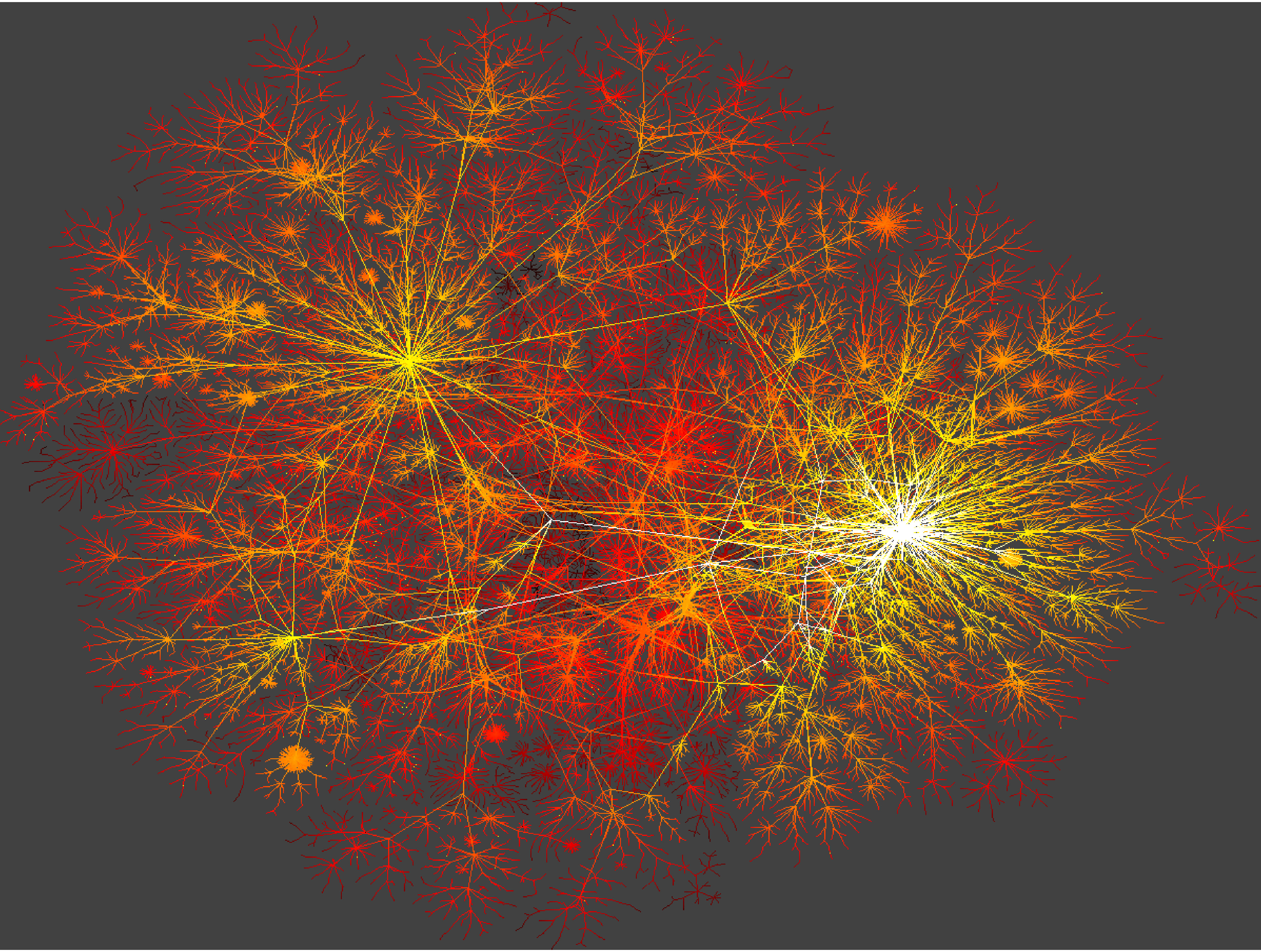
Mapping projects:

- Multi-probe reconstruction (router-level): traceroute
- Use of BGP tables for the Autonomous System level (domains)

- CAIDA, NLANR, RIPE, IPM, PingER, DIMES

Topology and performance measurements

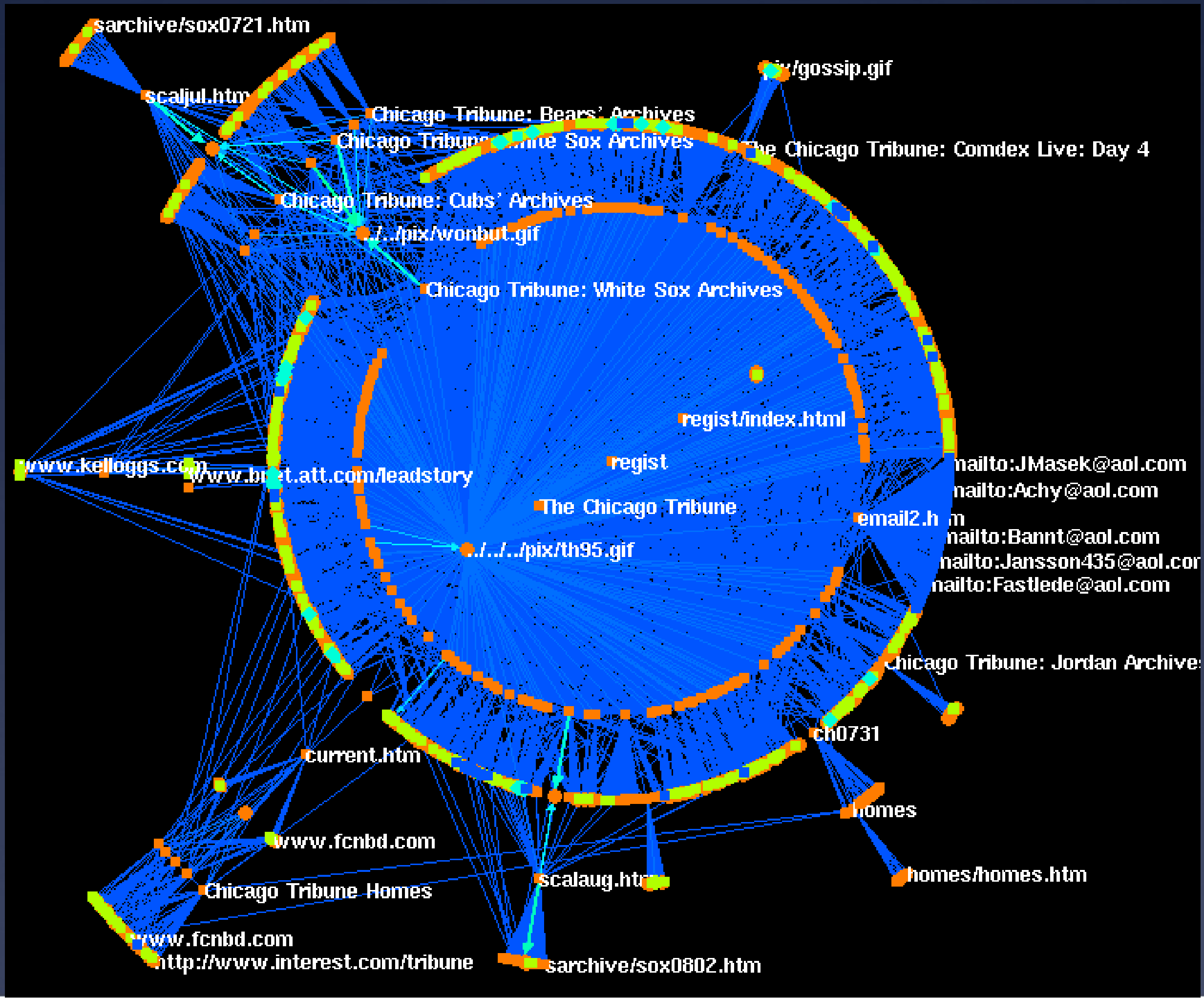

Netname:
(1717)
as-ebone(3215)
as-telianetse(3301)
bbn/gte(1)
digex(2548)
ebone(3269)
janet(786)
mci(3561)
sprint(1239)
uunet(701)



The World-Wide-Web

Virtual network to find and share information

- **web pages**
- **hyperlinks**



CRAWLS

Sampling issues

- social networks: various samplings/networks
- transportation network: reliable data
- biological networks: incomplete samplings
- Internet: various (incomplete) mapping processes
- WWW: regular crawls
- …

possibility of introducing biases in the measured network characteristics

Networks characteristics

Networks: of very different origins

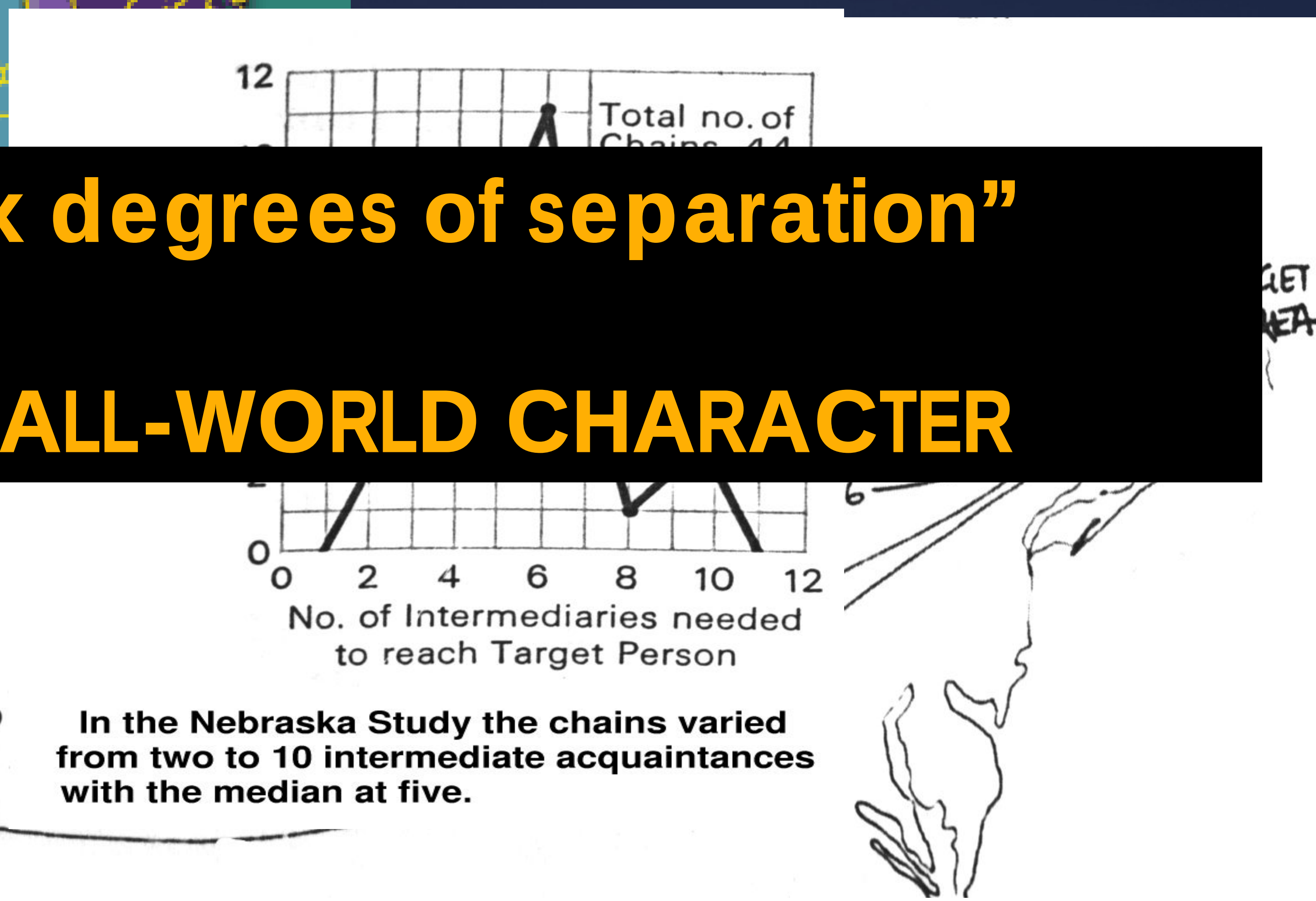
Do they have anything in common?
Possibility to find common properties?

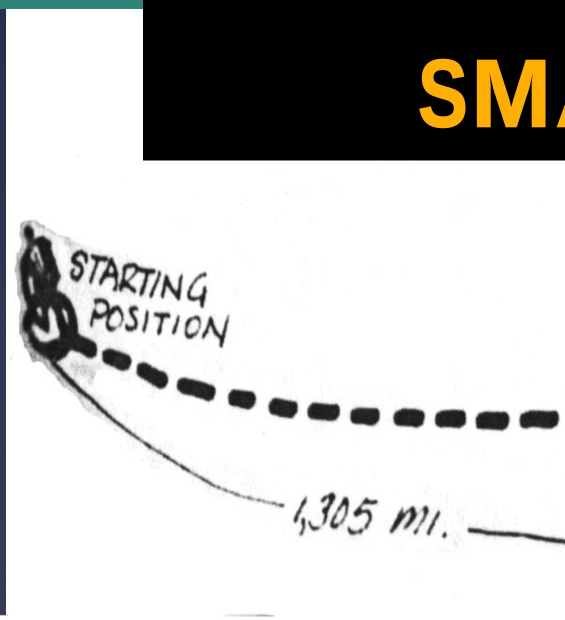
the abstract character of the graph representation
and graph theory allow to answer....

Social networks: Milgram’s experiment

Milgram, *Psych Today* **2**, 60 (1967)

Dodds et al., *Science* **301**, 827 (2003)



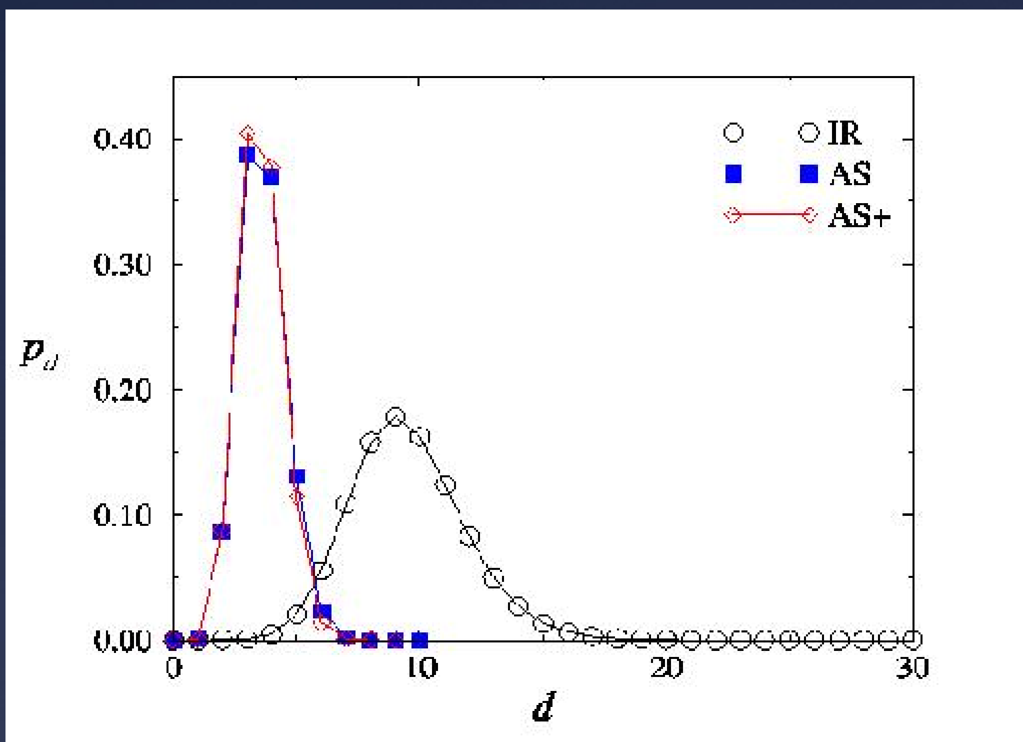


“Six degrees of separation”

SMALL-WORLD CHARACTER

Small-world properties

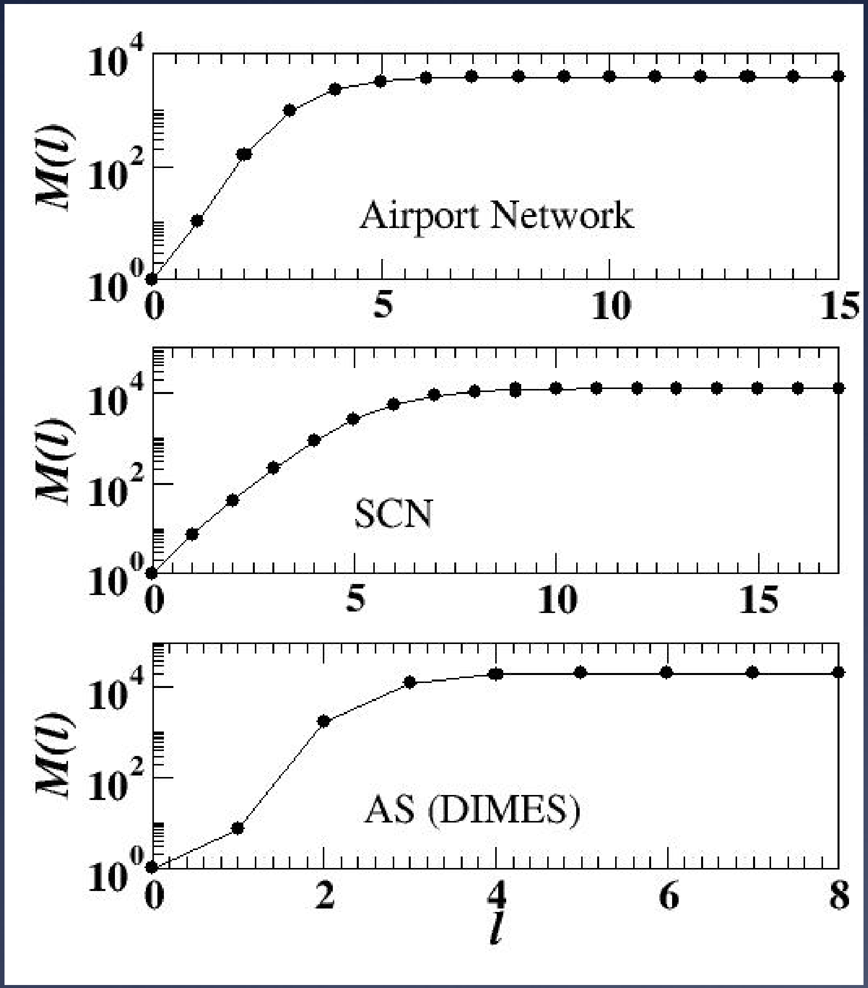
Internet:



Distribution of distances between two vertices

Small-world properties

Average number of vertices within a distance l



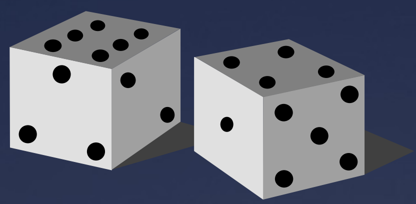
Airport network

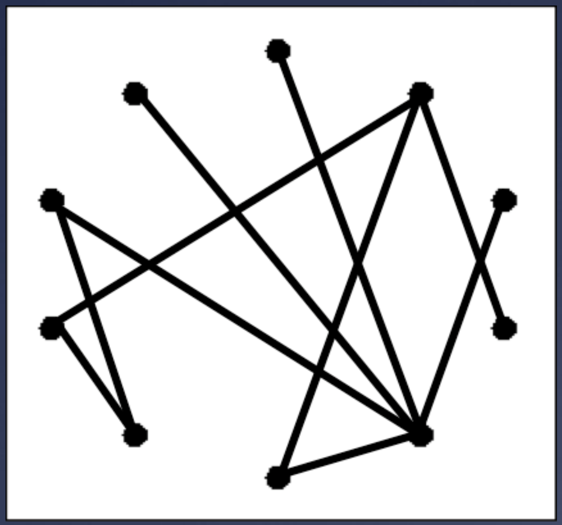
Scientific collaborations

Internet

Small-world properties

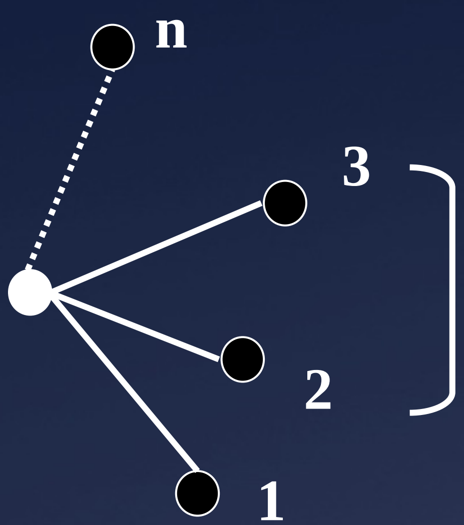
N vertices, edges with probability p:
static random graphs





short distances
(log N)

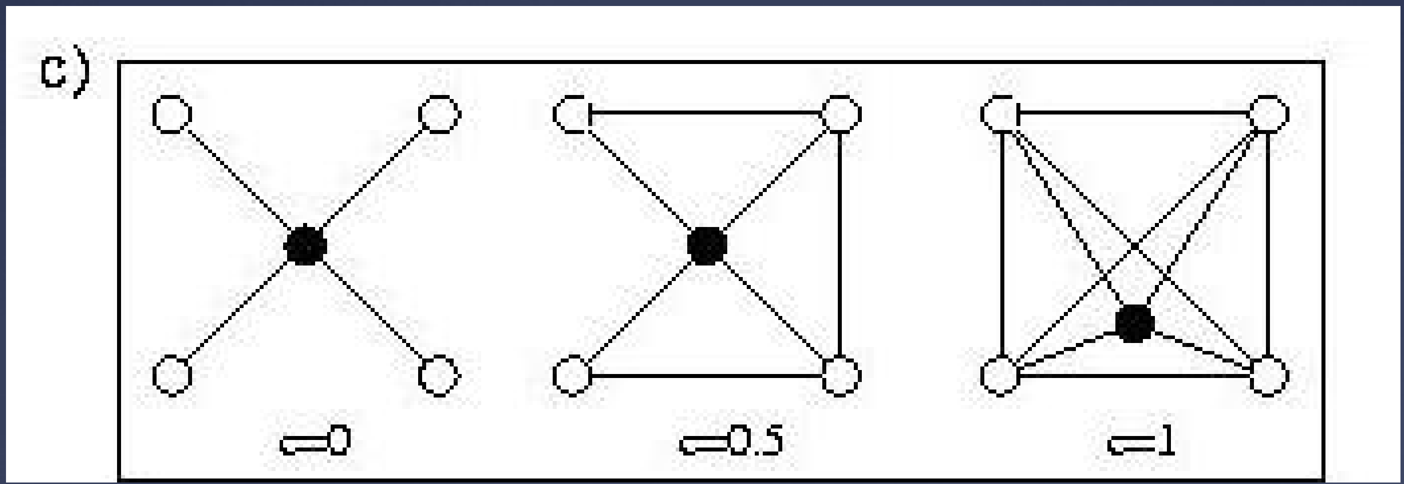
Clustering coefficient



Empirically: large clustering coefficients

Higher probability to be connected

Clustering: My friends will know each other with high probability (typical example: social networks)



Topological heterogeneity

Statistical analysis of centrality measures:

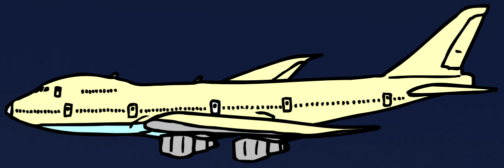
$P(k)=n_k/n=$probability that a randomly chosen vertex has degree k

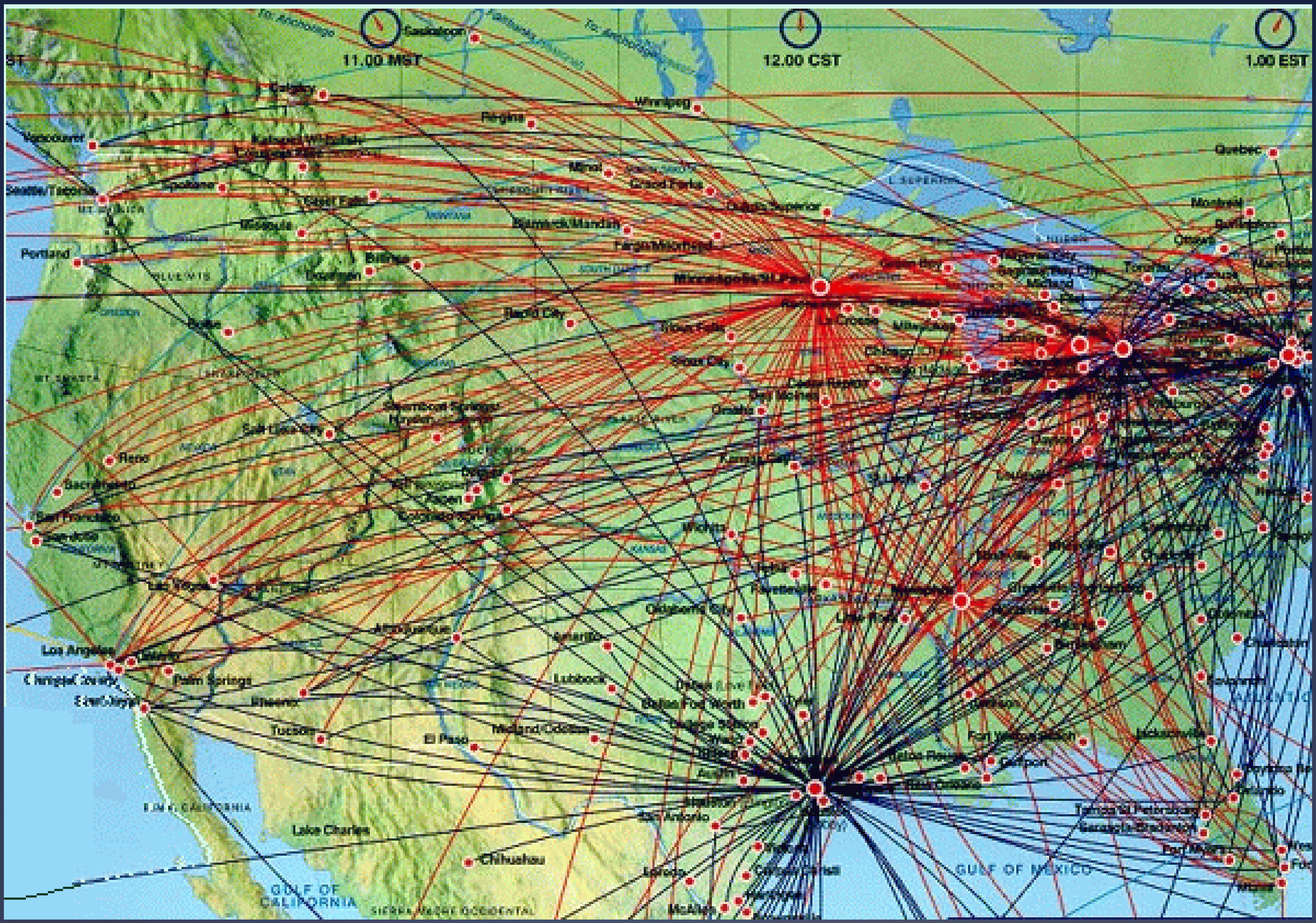
also: P(b), P(w)....

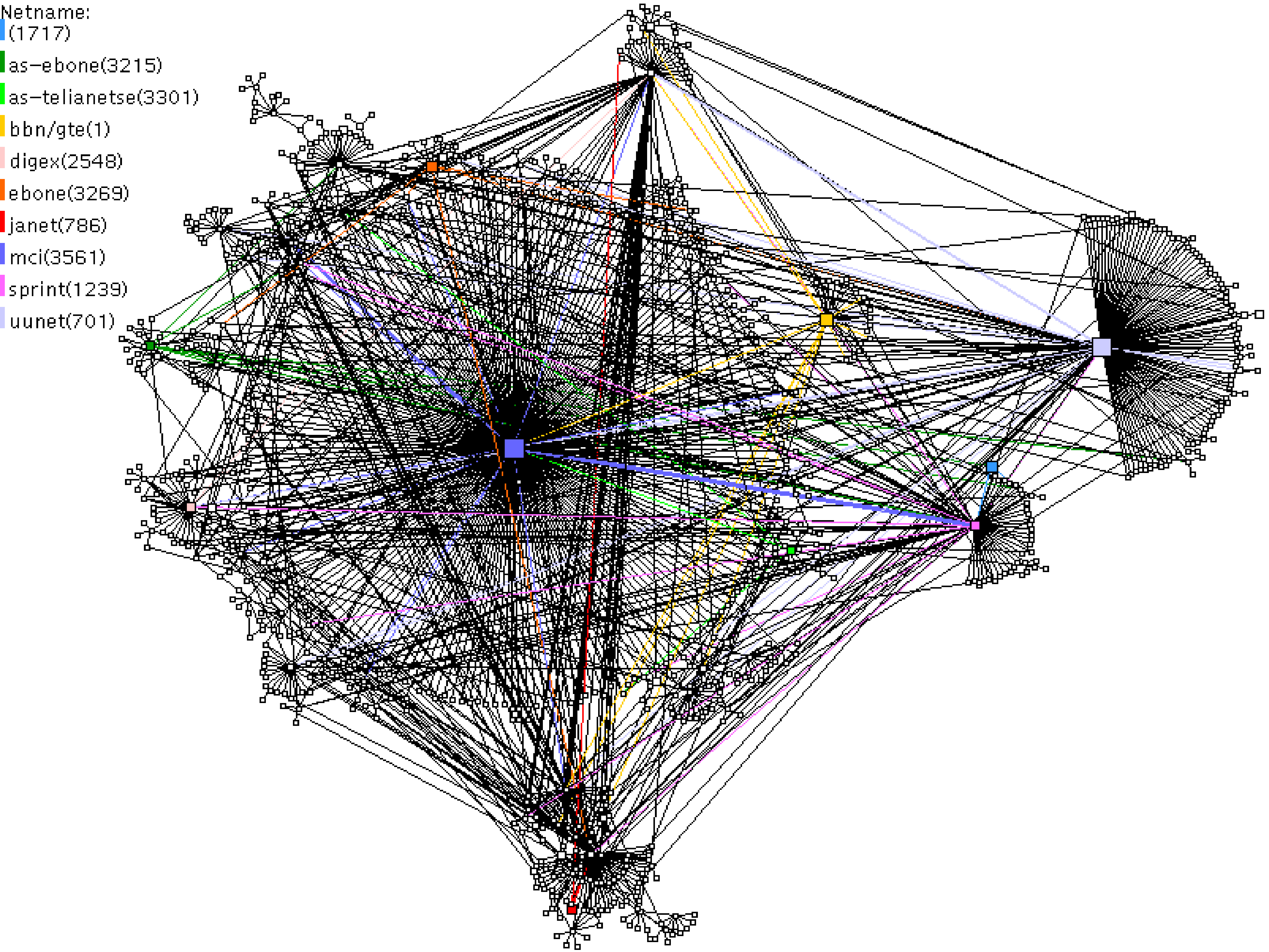
Two broad classes

- homogeneous networks: light tails
- heterogeneous networks: skewed, heavy tails

Airplane route network

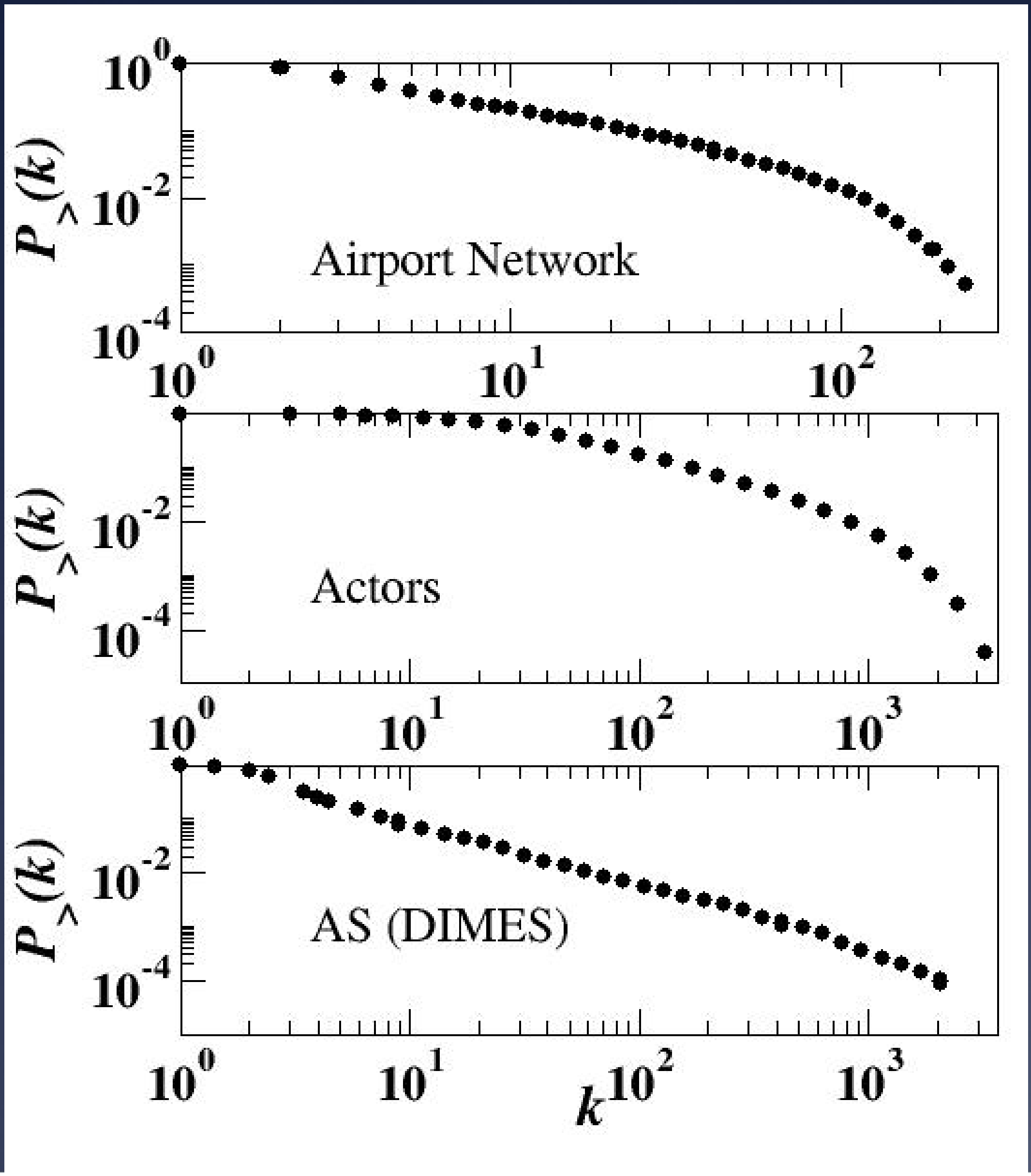





Netname:
(1717)
as-ebone(3215)
as-telianetse(3301)
bbn/gte(1)
digex(2548)
ebone(3269)
janet(786)
mci(3561)
sprint(1239)
uunet(701)

Topological heterogeneity

Statistical analysis of centrality measures

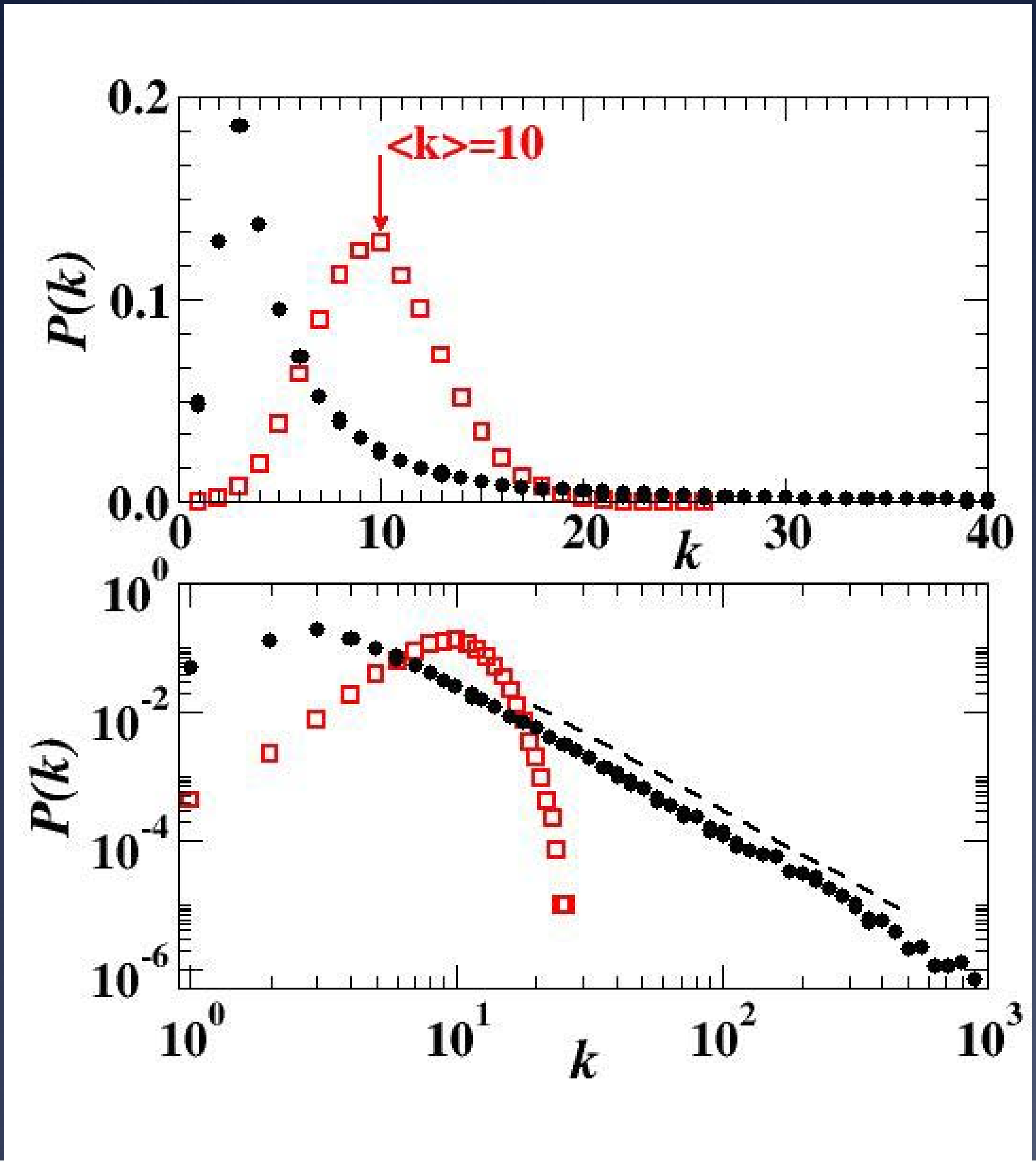


Broad degree distributions

Power-law tails
$P(k) \sim k^{-\gamma}$,
typically $2 < \gamma < 3$

Topological heterogeneity

Statistical analysis of centrality measures



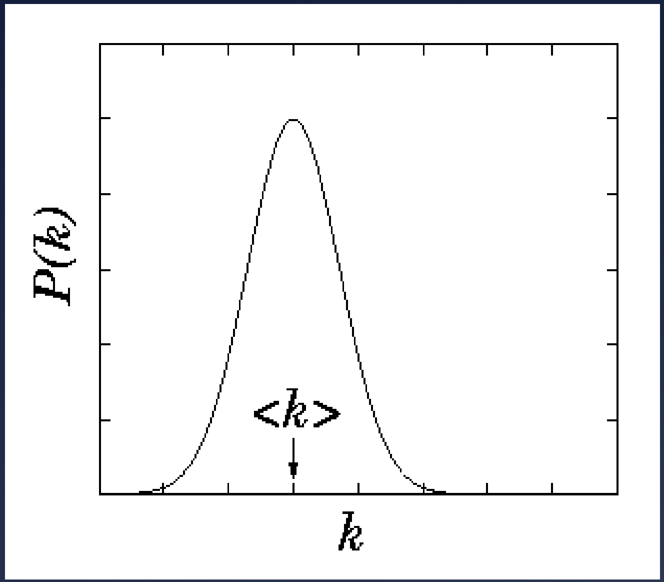
linear scale

Poisson
vs.
Power-law

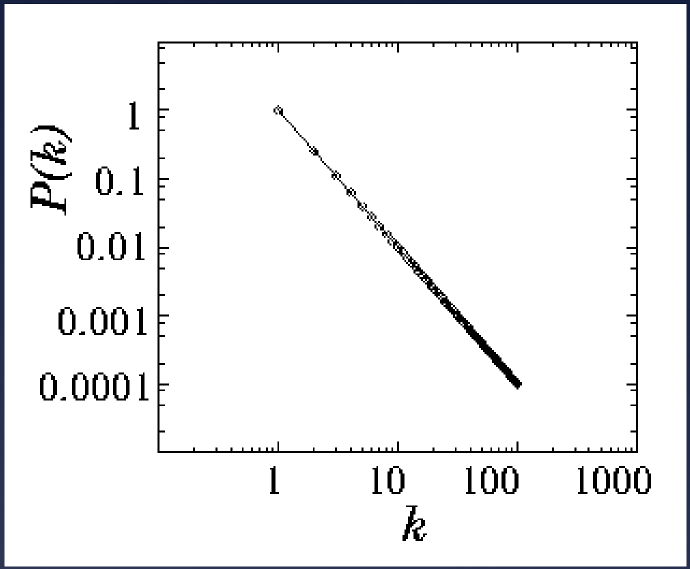
log-scale

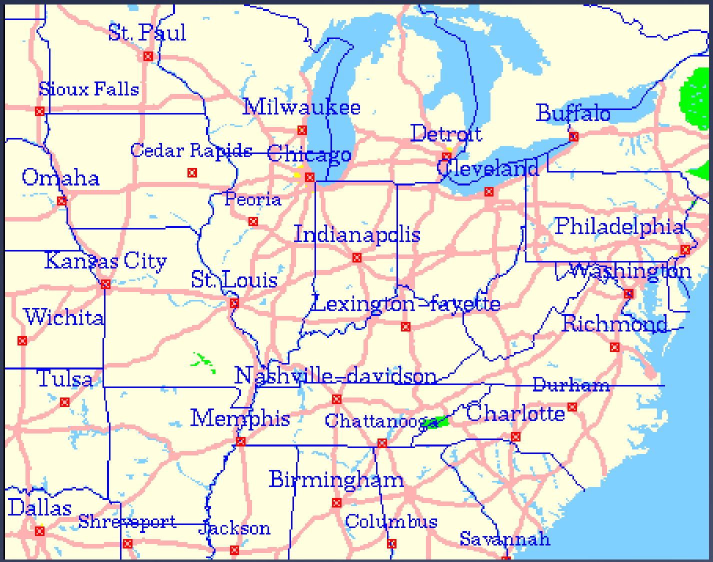
Exp. vs. Scale-Free

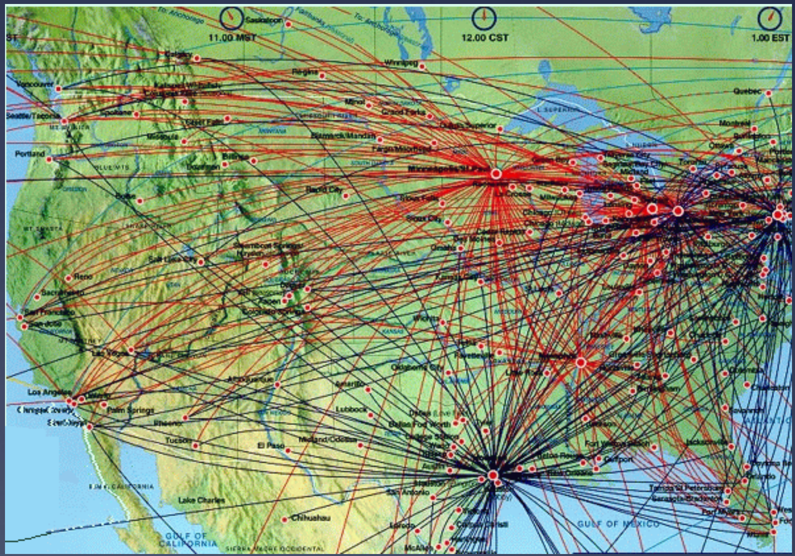
Poisson distribution



Power-law distribution







Exponential Network

Scale-free Network

Consequences

Power-law tails: $P(k) \sim k^{-\gamma}$

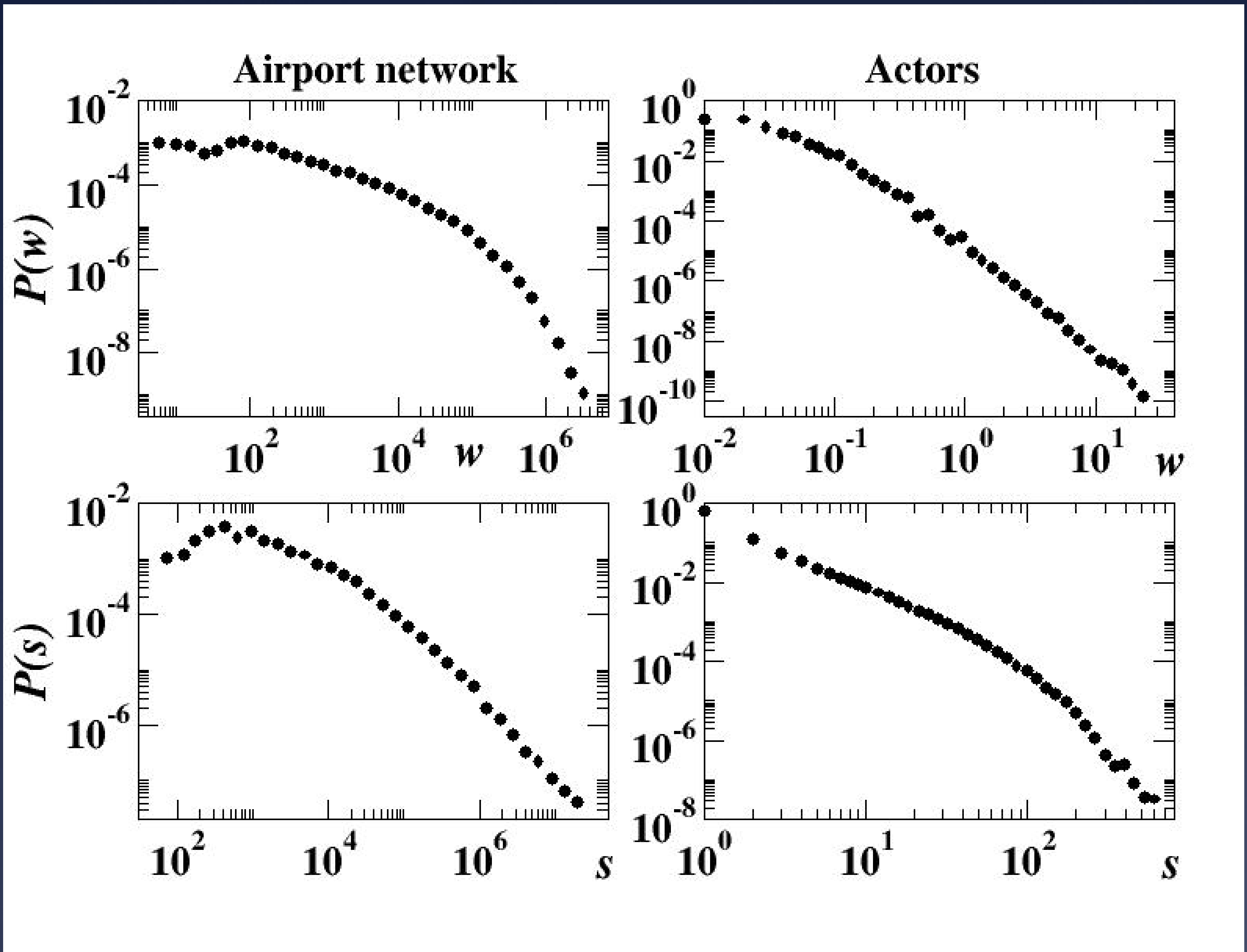
Averag
e $\langle k \rangle = \int kP(k)dk$

Fluctuations $\langle k^2 \rangle = \int k^2 P(k)dk \sim k_c^{3-\gamma}$

k_c=cut-off due to finite-size
diverging degree fluctuations for $\gamma < 3$

Level of heterogeneity: $\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}$

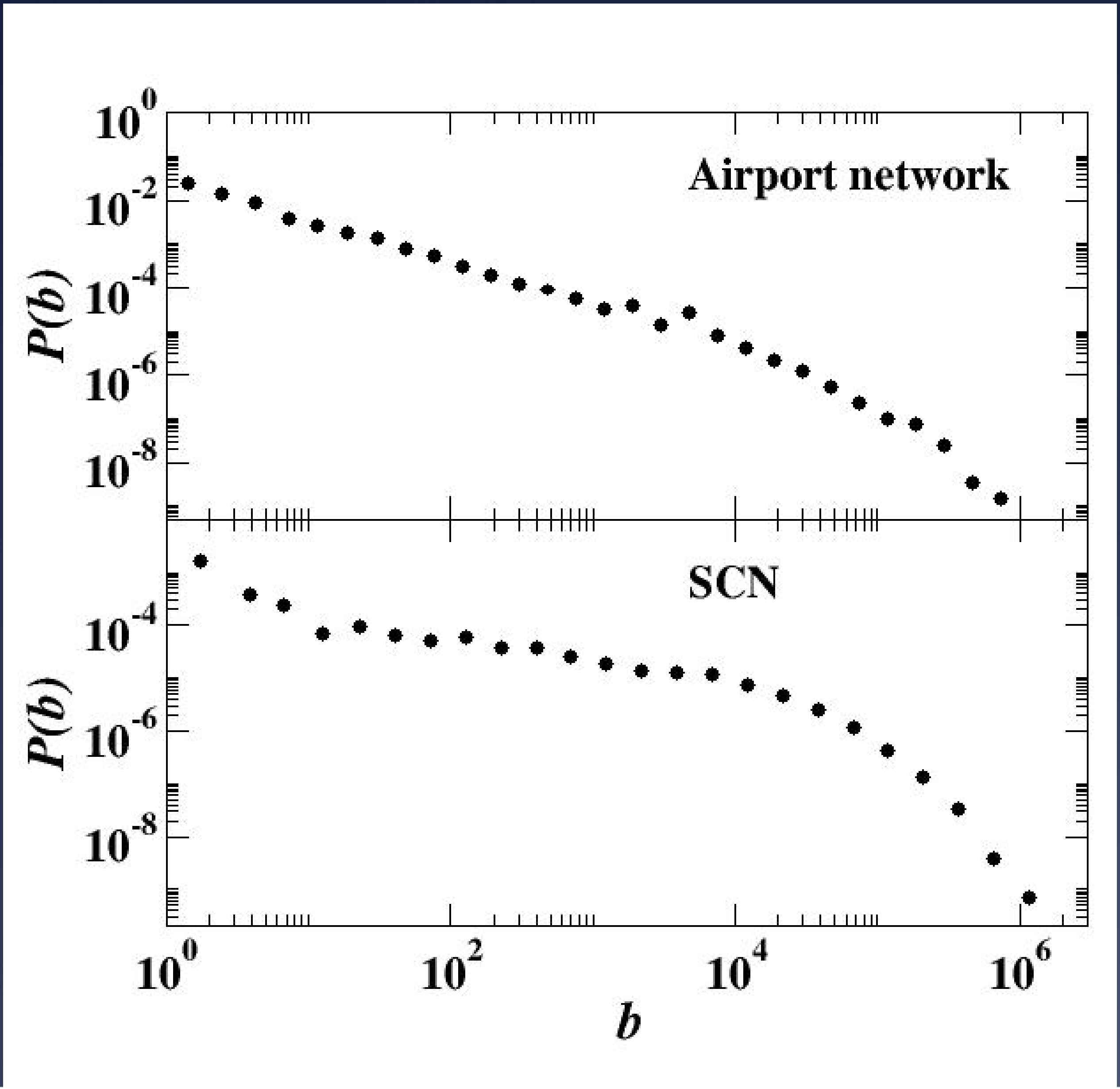
Other heterogeneity levels



Weights

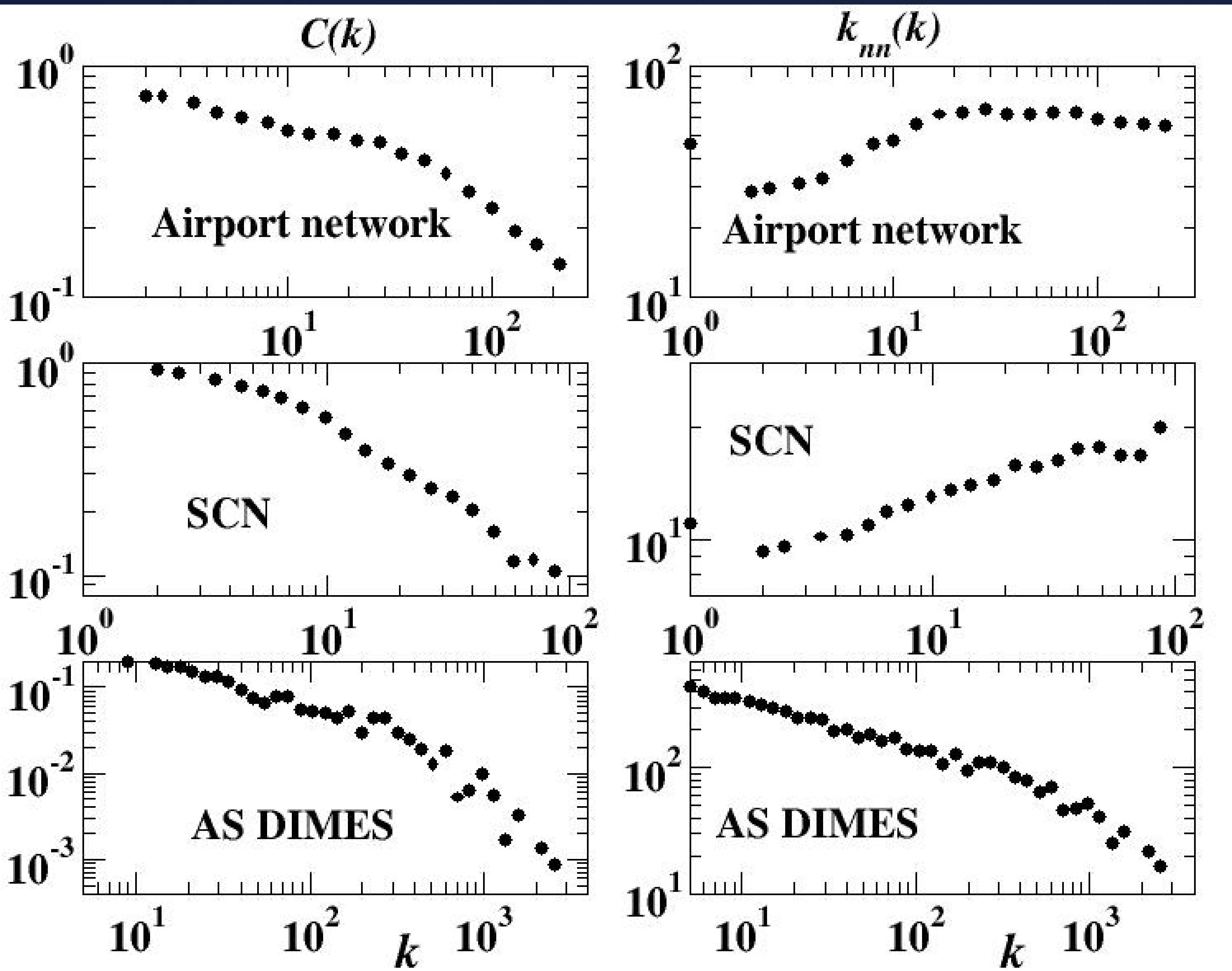
Strengths

Other heterogeneity levels



Betweenness centrality

Clustering and correlations



non-trivial structures

Real networks: summary!

	Network	Type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
Social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	[20, 415]
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	[105, 322]
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	[107, 181]
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	[310, 312]
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	[310, 312]
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				[8, 9]
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		[136]
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	[320]
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	[45]
	sexual contacts	undirected	2 810				3.2				[264, 265]
Information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	[14, 34]
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				[74]
	citation network	directed	783 339	6 716 198	8.57		3.0/–				[350]
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	[243]
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		[119, 157]
Technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	[86, 148]
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	[415]
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	[365]
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	[317]
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	[394]
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	[155]
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	[6, 353]
Biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	[213]
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	[211]
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	[203]
	freshwater food web	directed	92	997	10.84	1.90	–	0.40	0.48	−0.326	[271]
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	[415, 420]

Complex networks

Complex is not just "complicated"

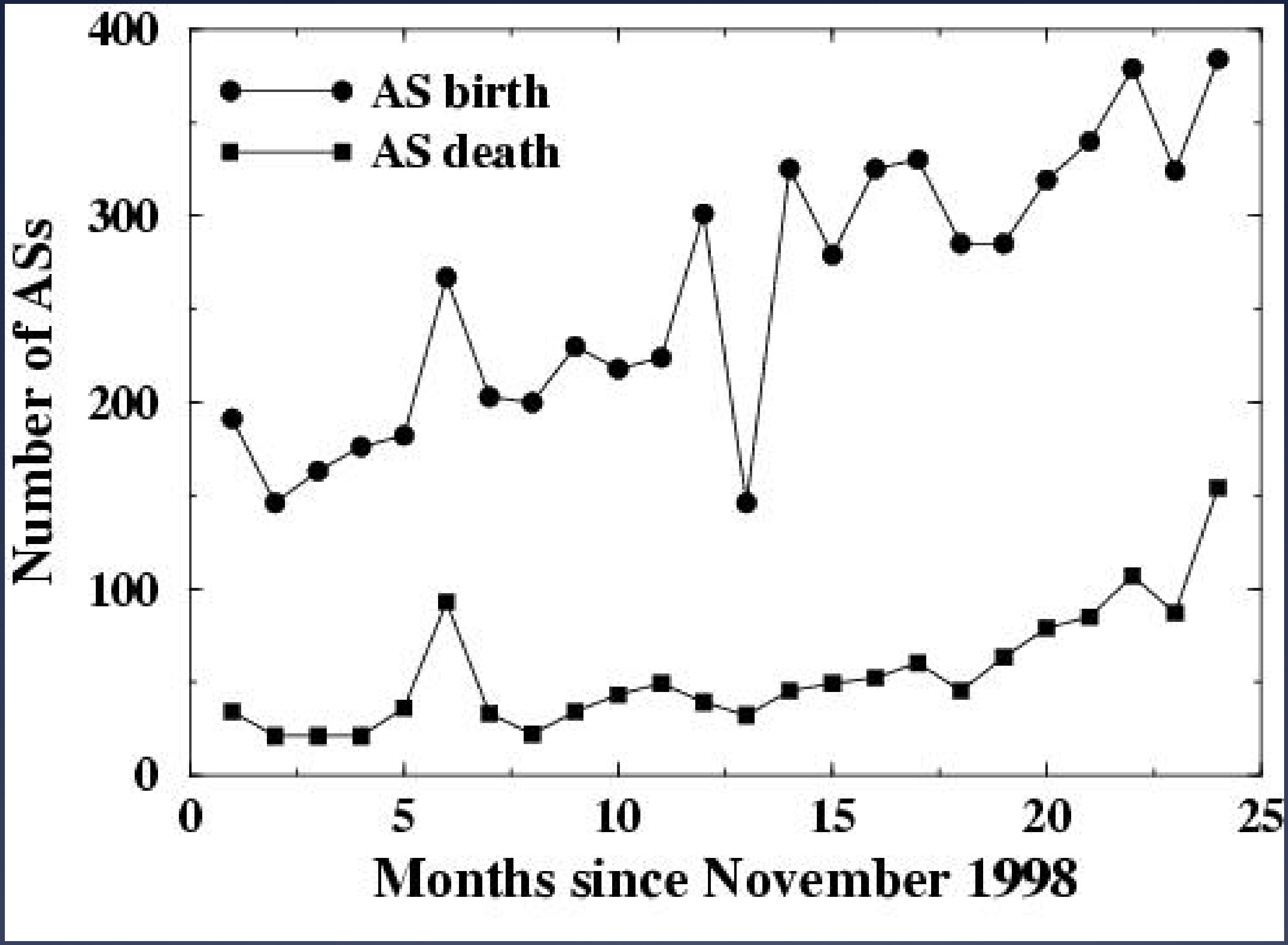
Cars, airplanes…=> complicated, not complex

Complex (no unique definition):

- many interacting units
- no centralized authority, self-organized
- complicated at all scales
- evolving structures
- emerging properties (heavy-tails, hierarchies…)

Examples: Internet, WWW, Social nets, etc…

Example: Internet growth



Main features of complex networks

- Many interacting units
- Self-organization
- Small-world
- Scale-free heterogeneity
- Dynamical evolution

Standard graph theory

Random graphs

- Static
- Ad-hoc topology

Example: Internet topology generators
Modeling of the Internet structure with ad-hoc algorithms tailored on the properties we consider more relevant

Statistical physics approach

Microscopic processes of the many component units

↓

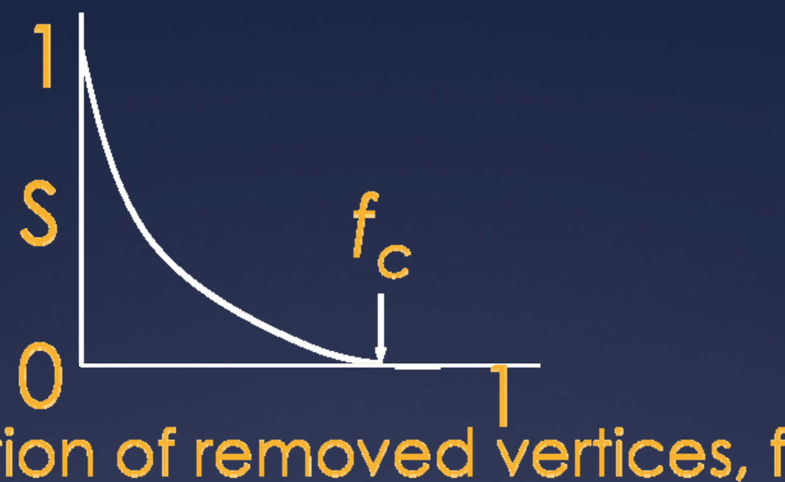
Macroscopic statistical and dynamical properties of the system

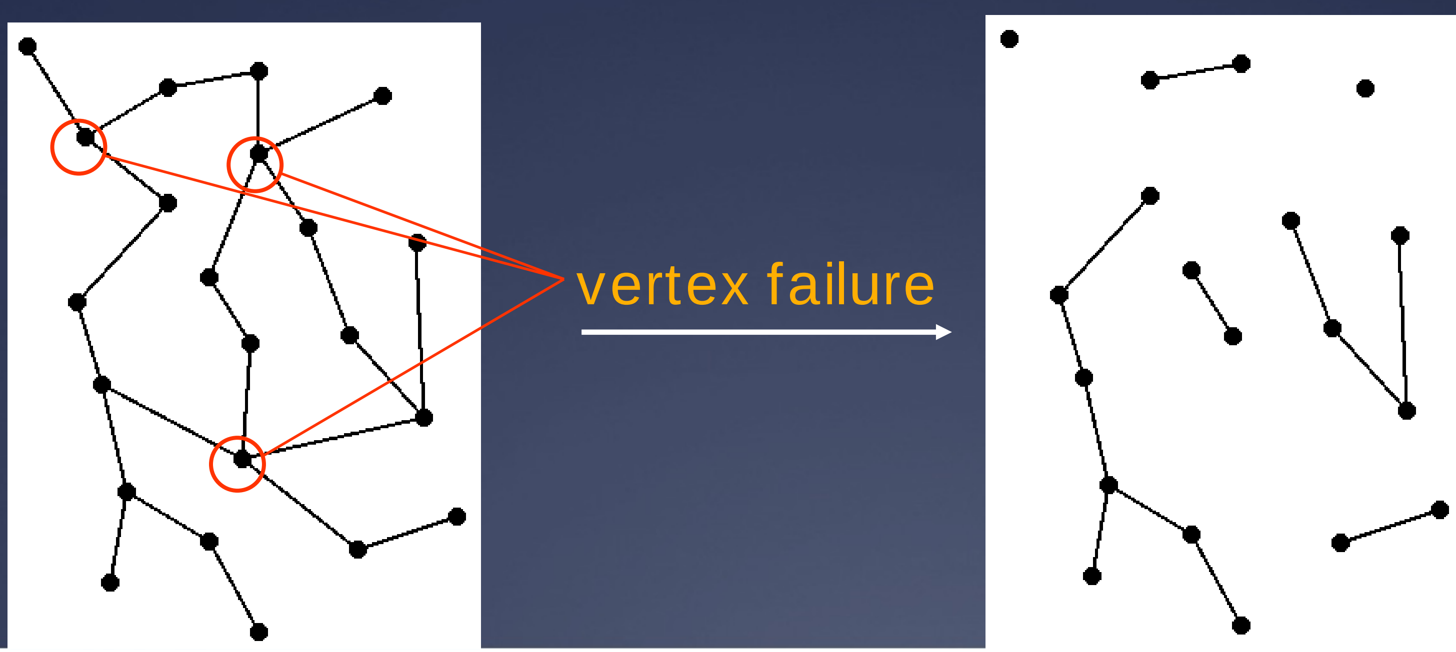
Cooperative phenomena Complex topology ⟷ Natural outcome of the dynamical evolution

Robustness

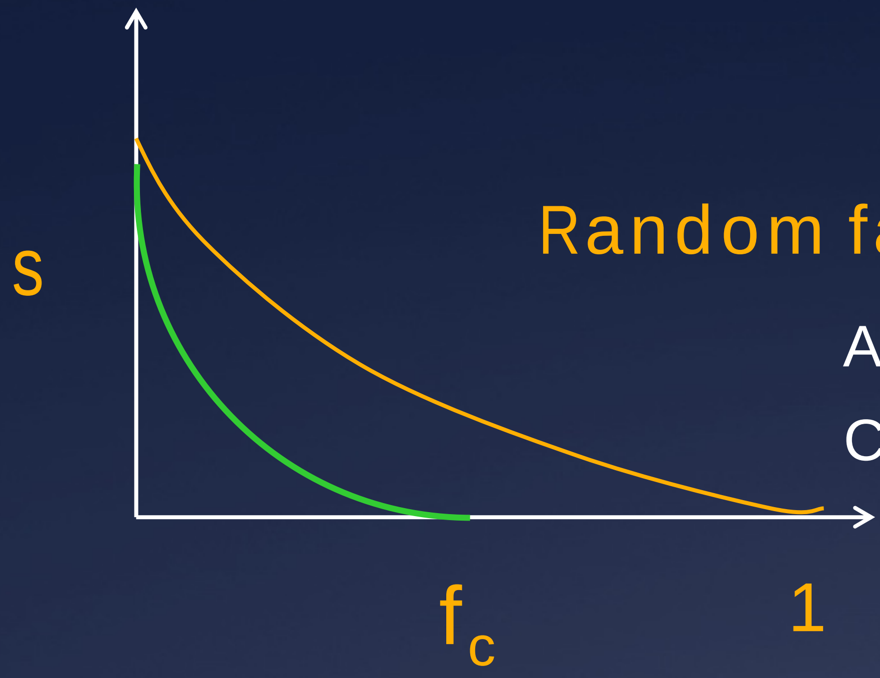
Complex systems maintain their basic functions even under errors and failures
(cell → mutations; Internet → router breakdowns)

S: fraction of giant component





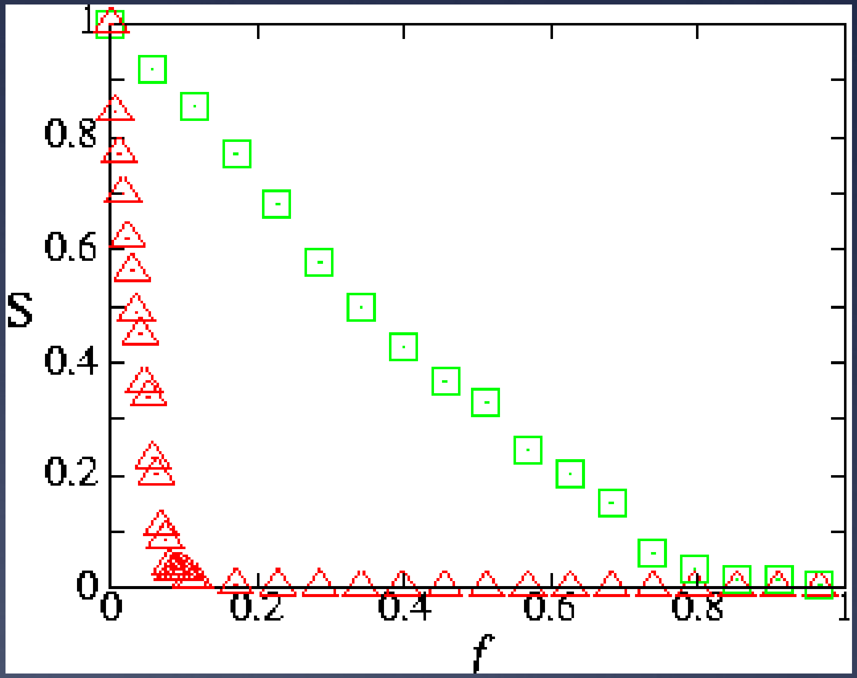
Case of Scale-free Networks



Random failure $f_c = 1$ $(2 < \gamma \leq 3)$

Attack = progressive failure of the most Connected vertices $f_c < 1$

Internet maps

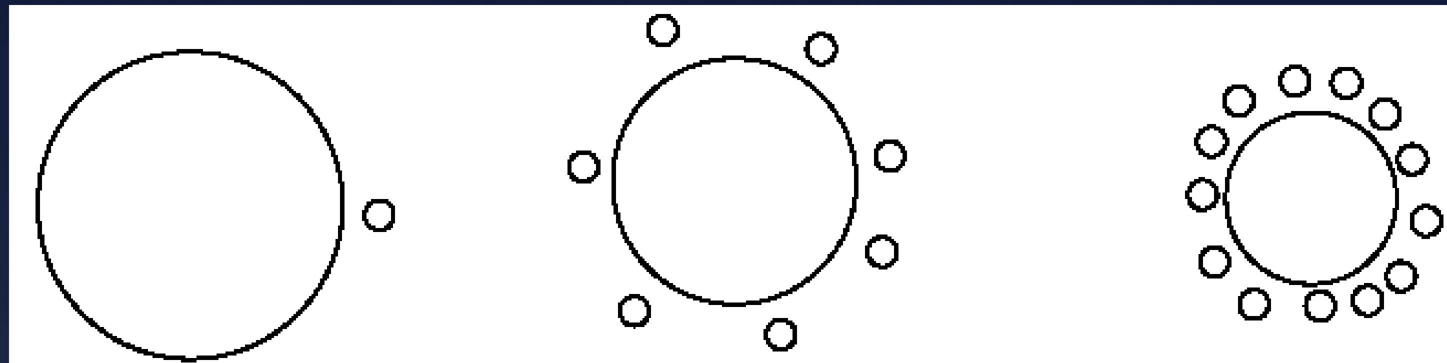


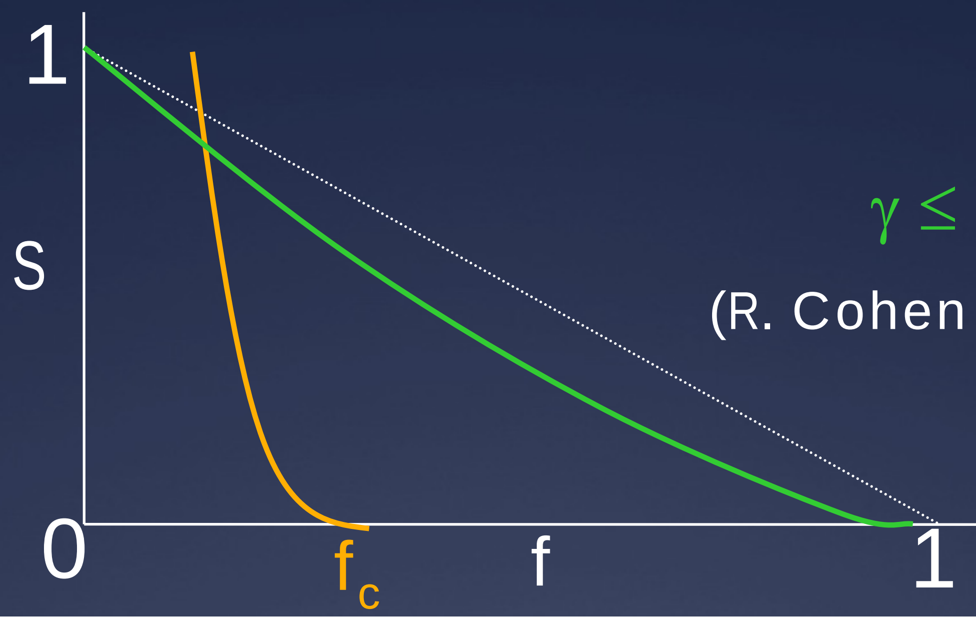
R. Albert, H. Jeong, A.L. Barabasi, Nature 406 378 (2000)

Failures vs. attacks

Failures

Topological error tolerance

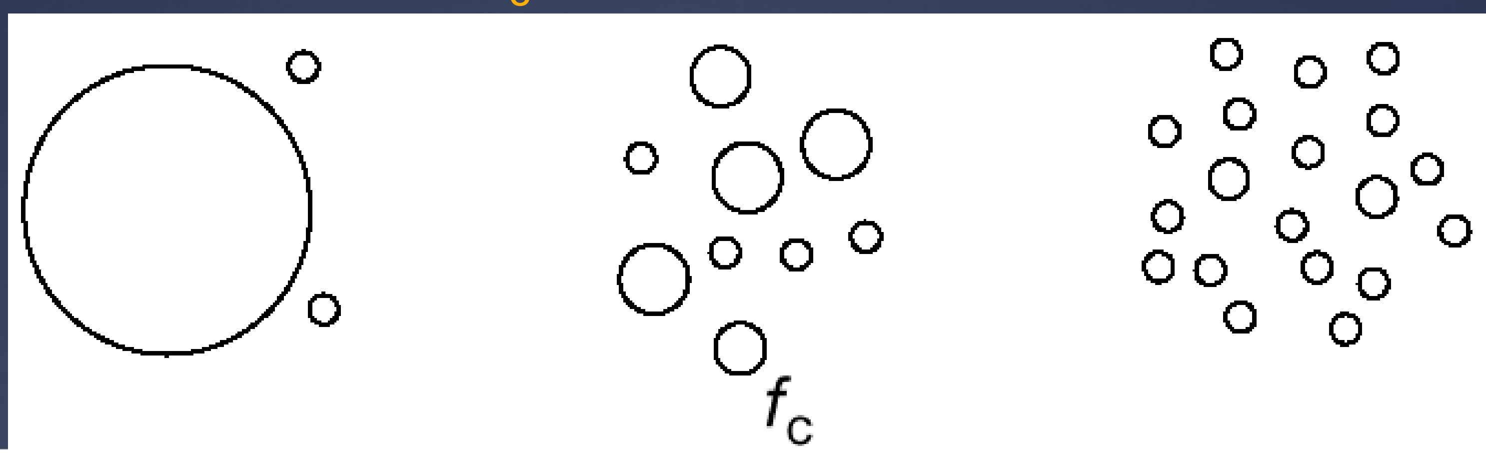




$\gamma \leq 3 : f_c = 1$

(R. Cohen et al PRL, 2000)

Attacks



Failures = percolation

f=fraction of vertices removed because of failure

p=1-f

p=probability of a vertex to be present in a percolation problem

Question: existence or not of a giant/percolating cluster, i.e. of a connected cluster of nodes of size O(N)

Betweenness

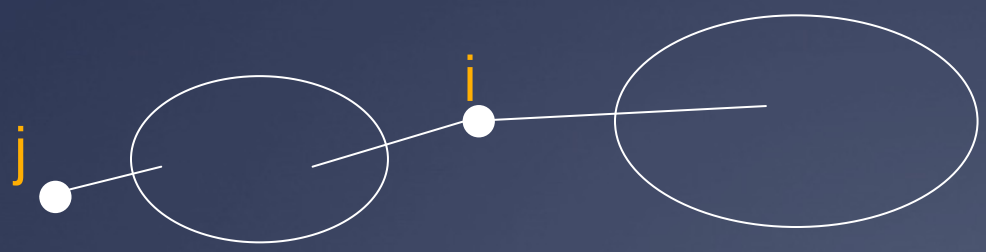
⇒ measures the “centrality” of a vertex i:

for each pair of vertices (l,m) in the graph, there are

σ^{lm} shortest paths between l and m

σ_i^{lm} shortest paths going through i

b_i is the sum of $\sigma_i^{lm} / \sigma^{lm}$ over all pairs (l,m)



b_i is large

b_j is small

Attacks: other strategies

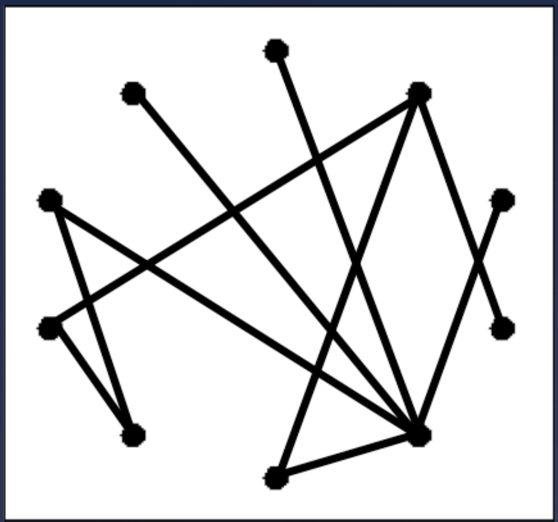
* Most connected vertices
* Vertices with largest betweenness
* Removal of edges linked to vertices with large k
* Removal of edges with largest betweenness
* Cascades
* ...

P. Holme et al (2002); A. Motter et al. (2002);
D. Watts, PNAS (2002); Dall'Asta et al. (2006)...

Classical random graphs

Solomonoff & Rapoport (1951), Erdös-Rényi (1959)

$\mathcal{G}_{n,p} =$ n vertices, edges with probability p: static random graphs



Average number of edges: $\langle E \rangle = pn(n-1)/2$

Average degree: $\langle k \rangle = p(n-1)$

p=c/n to have finite average degree

Related formulation

$\mathcal{G}_{n,m} =$ n vertices, m edges: each configuration is equally probable

Classical random graphs

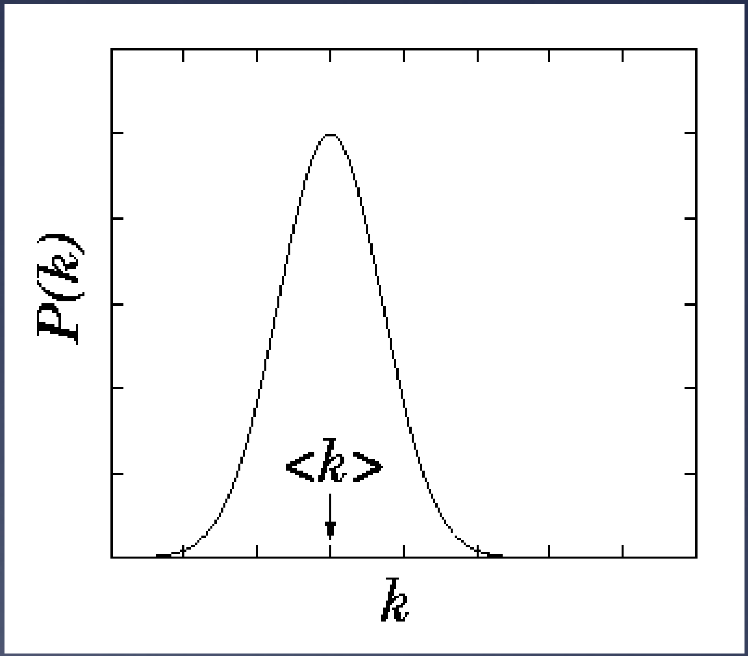
Probability to have a vertex of degree k

- connected to k vertices,
- not connected to the other n-k-1

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-k-1}$$

Large N, fixed pN=< k > : Poisson distribution

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



Exponential decay at large k

Classical random graphs

Properties

I. Poisson degree distribution (on large graphs)

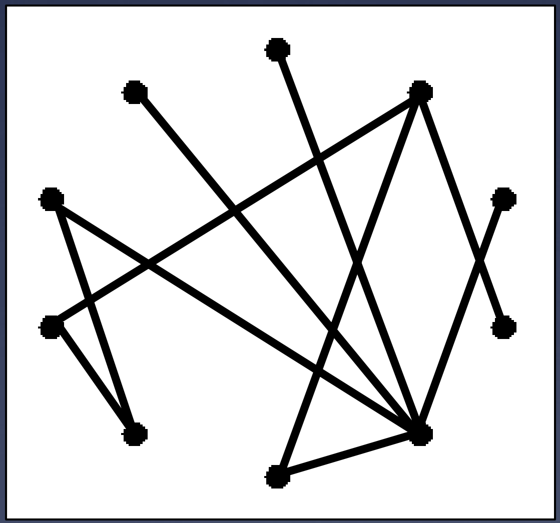
II. Small clustering coefficient: $\langle C \rangle = p = \langle k \rangle / n$ (goes to zero in the limit of infinite graph size for sparse graphs)

III. Short distances: diameter $l \sim \log(n)/\log(\langle k \rangle)$ (number of neighbors at distance d: $\langle k \rangle^d$

Classical random graphs

<k> < 1: many small subgraphs

< k > > 1: giant component + small subgraphs



Classical random graphs

Self-consistent equation for relative size of giant component

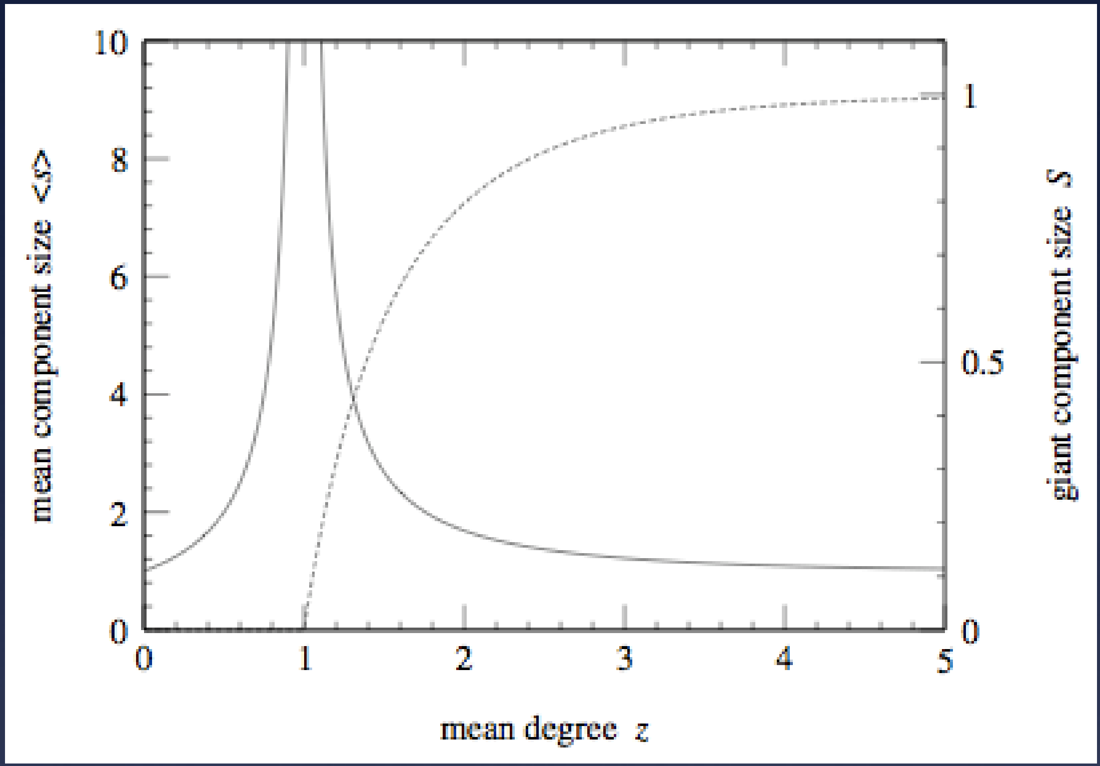
$$u = \sum_{k=0}^{\infty} P(k) u^k = e^{-<k>} \sum_{k=0}^{\infty} \frac{(< k > u)^k}{k!} = e^{<k>(u-1)}$$

$$S = 1 - u \quad \rightarrow \quad S = 1 - e^{-<k>S}$$

Average cluster size

$$\langle s \rangle = \frac{1}{1 - < k > + < k > S}$$

Classical random graphs



Near the transition (<k>~1)

$$S \sim (<k> -1)^{\beta}$$

$$<s> \sim |<k> -1|^{-\gamma}$$

$$\beta = 1, \gamma = 1$$

Generalized random graphs

The configuration model (Molloy & Reed, 1995, 1998)

Basic idea: building random graphs with arbitrary degree distributions

How it works

I. Choose a degree sequence compatible with some given distribution
II. Assign to each vertex his degree, taken from the sequence, in that each vertex has as many outgoing stubs as its degree
III. Join the stubs in random pairs, until all stubs are joined

Generalized random graphs

Simple model: easy to handle analytically

Important point: the probability that the degree of vertex reached by following a randomly chosen edge is k is not given by P(k)!

Reason: vertices with large degree have more edges and can be reached more easily than vertices with low degree

Conclusion: distribution of degree of vertices at the end of randomly chosen edge is proportional to kP(k)

Generalized random graphs

Excess degree: number of edges leaving a vertex reached from a randomly selected edge, other than the edge followed

Distribution of excess degree: $Q(k)$

$$Q(k) = \frac{(k+1)P(k+1)}{\sum_k kP(k)} = \frac{(k+1)P(k+1)}{< k >}$$

Generalized random graphs

Chance of finding loops within small (i.e. non-giant) components goes as $O(n^{-1})$

Consequence: graphs of the configuration model are essentially loopless, tree-like, unlike real-world networks

Generating functions

$$G_0(x) = \sum_{k=0}^{\infty} P(k) x^k$$

$$G_1(x) = \sum_{k=0}^{\infty} Q(k) x^k$$

Generalized random graphs

$$G_1(x) = G_0'(x) / < k >$$

$$< k >= G_0'(1)$$

$$< k^2 > - < k >= G_0'(1) G_1'(1)$$

$$< s >= 1 + \frac{< k >^2}{2 < k > - < k^2 >}$$

Condition for existence of giant component:

$$\sum_k k(k-2) P(k) = 0 \quad \Longleftrightarrow \quad G_1'(1) = 1$$

Generalized random graphs

$u =$ probability that a randomly chosen edge is not in the giant component

$$u = \sum_{k=0}^{\infty} Q(k) u^k = G_1(u)$$

$$1 - S = \sum_{k=0}^{\infty} P(k) u^k = G_0(u)$$

Self-consistent equations for the relative size S of the giant component

Generalized random graphs

Example: power-law degree distribution

$$P(k) = \begin{cases} 0 & \text{for } k = 0 \\ k^{-\alpha}/\zeta(\alpha) & \text{for } k \geq 1 \end{cases}$$

$$G_0(x) = \frac{Li_\alpha(x)}{\zeta(\alpha)}, \qquad G_1(x) = \frac{Li_{\alpha-1}(x)}{x\zeta(\alpha-1)}$$

$$\mathrm{Li}_s(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^s}$$

$$G_1'(1) = 1 \quad \rightarrow \quad \zeta(\alpha-2) = 2\zeta(\alpha-1)$$

Critical exponent value: $\alpha_c = 3.4788$

For $\alpha < \alpha_c$ there is always a giant connected component!

For $\alpha > \alpha_c$ there is no giant connected component!

Generalized random graphs

$$S = 1 - G_0(u), \quad u = G_1(u)$$

$$u = \frac{\mathrm{Li}_{\alpha-1}(u)}{u\zeta(\alpha-1)}$$

For $\alpha < 2$, u=0 and S=1 ➔ All vertices are in the giant component!

W. Aiello, F. Chung, L. Lu, *Proc. 32th ACM Symposium on Theory of Computing* 171-180 (2000)

Plan of the course

I. Networks: definitions, characteristics, basic concepts in graph theory
II. Real world networks: basic properties. Models I
III. Models II
IV. Community structure I
V. Community structure II
VI. Dynamic processes in networks